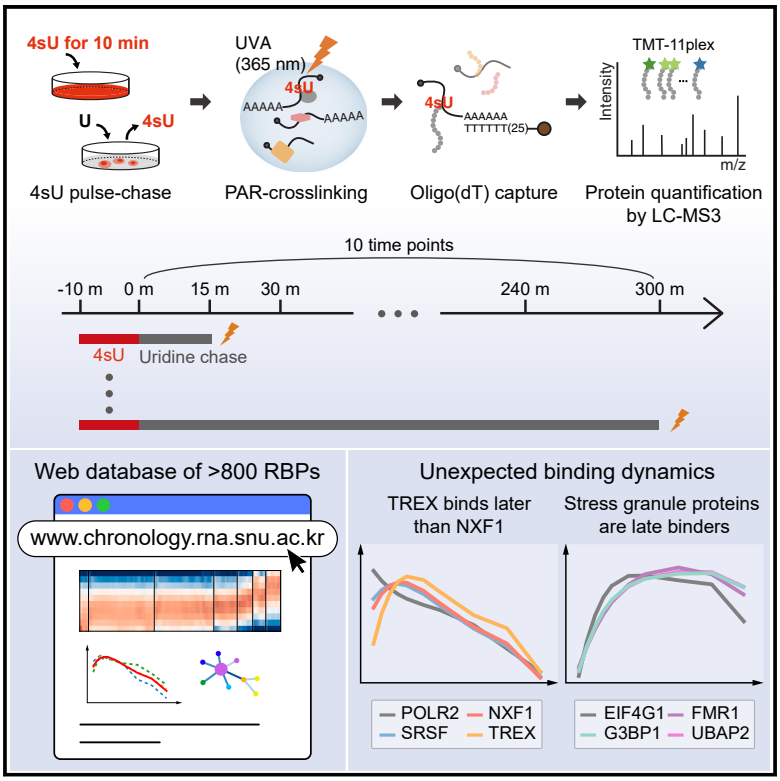
Molecular Cell

Resource

Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle

Graphical abstract



Highlights

- Temporal mRNA interactome data reveal mRNP remodeling throughout the life cycle
- Overall, the sequential order of mRNA binding aligns well with known functions
- Some RBPs including TREX and stress granule components show unexpected dynamics
- Information on over 800 human RBPs is accessible at chronology.rna.snu.ac.kr

Authors

Yeon Choi, Buyeon Um, Yongwoo Na, Jeesoo Kim, Jong-Seo Kim, V. Narry Kim

Correspondence

jongseokim@snu.ac.kr (J.-S.K.), narrykim@snu.ac.kr (V.N.K.)

In brief

Choi et al. conducted mRNA-interactome-capture experiments in a pulse-chase manner, revealing the timing of mRNA-protein interactions throughout the mRNA life cycle. Unexpected binding dynamics were observed in some RNA binding proteins (RBPs), including NXF1 and the TREX components. Information on over 800 human RBPs is accessible via a website.

Choi et al., 2024, Molecular Cell *84*, 1–19
May 2, 2024
https://doi.org/10.1016/j.molcel.2024.03.012

CellPress

Molecular Cell

CellPress

Resource

Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle

Yeon Choi,[1,2,3] Buyeon Um,[1,2,3] Yongwoo Na,[1,2] Jeesoo Kim,[1,2] Jong-Seo Kim,[1,2,*] and V. Narry Kim[1,2,4,*]
[1]Center for RNA Research, Institute for Basic Science, Seoul 08826, Republic of Korea
[2]School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea
[3]These authors contributed equally
[4]Lead contact
*Correspondence: jongseokim@snu.ac.kr (J.-S.K.), narrykim@snu.ac.kr (V.N.K.)
https://doi.org/10.1016/j.molcel.2024.03.012

SUMMARY

mRNAs continually change their protein partners throughout their lifetimes, yet our understanding of mRNA-protein complex (mRNP) remodeling is limited by a lack of temporal data. Here, we present time-resolved mRNA interactome data by performing pulse metabolic labeling with photoactivatable ribonucleoside in human cells, UVA crosslinking, poly(A)+ RNA isolation, and mass spectrometry. This longitudinal approach allowed the quantification of over 700 RNA binding proteins (RBPs) across ten time points. Overall, the sequential order of mRNA binding aligns well with known functions, subcellular locations, and molecular interactions. However, we also observed RBPs with unexpected dynamics: the transcription-export (TREX) complex recruited posttranscriptionally after nuclear export factor 1 (NXF1) binding, challenging the current view of transcription-coupled mRNA export, and stress granule proteins prevalent in aged mRNPs, indicating roles in late stages of the mRNA life cycle. To systematically identify mRBPs with unknown functions, we employed machine learning to compare mRNA binding dynamics with Gene Ontology (GO) annotations. Our data can be explored at chronology.rna.snu.ac.kr.

INTRODUCTION

The life cycle of eukaryotic mRNA involves several distinct stages: transcription, precursor mRNA (pre-mRNA) processing, nuclear export, translation, and decay. In each stage, mRNAs interact with a specific set of RNA-binding proteins (RBPs) to form mRNA-protein complexes (mRNPs).[1,2] These RBPs govern the activity, localization, and stability of mRNA and influence its transition to the subsequent stage of the life cycle. Therefore, unveiling the stage-specific repertoire of RBPs is vital for understanding mRNA regulation.

Since the initial observation of mRNPs in the 1950s, many researchers have biochemically purified mRNPs and discovered core mRNP components, such as cap binding proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs), serine/arginine-rich (SR) splicing factors, and poly(A) binding proteins (PABPs).[1,3,4] More recently, studies have explored the interaction between mRNAs and RBPs comprehensively, employing high-throughput approaches such as the RNA interactome capture (RIC). In RIC experiments, RNA-protein partners are first crosslinked, followed by RNP enrichment, protein digestion, and liquid chromatography-tandem mass spectrometry (LC-MS/MS).[5–8] MS-based techniques collectively reported more than 6,000 human proteins as potential RBPs,[9] and it is believed that the human genome encodes at least 1,542 RBPs.[10] These methods have been applied to numerous biological contexts, revealing the diverse biological roles of RBPs.[11–14]

While the aforementioned approaches greatly advanced our knowledge of RNA-protein interactions, they provide unsynchronized mixed pools of mRNPs without temporal resolution. Since mRNPs involved in distinct stages of the mRNA life cycle (e.g., processing vs. translation) are expected to be fundamentally different from one another, there is a need to collect and analyze mRNPs specific to each stage. Time-resolved profiling would help reveal the compositional changes, offering a more comprehensive understanding of RBP functions. An earlier study examined nascent RNPs by using 5-ethynyluridine labeling followed by click chemistry-based RNA capture, but this approach mainly yielded proteins bound to abundant noncoding RNAs and had a limited number of time points (0.5, 2, and 16 h).[15]

In this study, we aimed to investigate mRNP remodeling over time by developing a time-resolved RIC technique. We enriched mRNAs of specific "ages," using pulse-chase metabolic labeling with 4-thio-uridine (4sU) and selective crosslinking under 365 nm light (ultraviolet light A [UVA]) across 10 time points.[16,17] This longitudinal analysis revealed the RNA binding dynamics of 734 mRNA binding proteins (mRBPs). We further integrated these temporal RNA interaction data with subcellular localization, RNP granule formation, protein-protein interaction (PPI), protein-mRNA interaction, viral RNA interaction, and Gene

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Ontology (GO) data to identify RBPs with as-yet-unknown functions and provided critical insights into mRNP remodeling. The chronological data of mRNA-protein interaction from this study can be accessed at: https://chronology.rna.snu.ac.kr/.

RESULTS

Time-resolved mRNA interactome profiling by 4sU pulse chasing

To identify RBPs that associate with mRNAs at specific time points following RNA synthesis, we employed a pulse-chase approach (Figures 1A and 1B). HeLa cells were metabolically labeled with 0.5 mM 4sU for 10 min, the shortest period that provided sufficient amounts of crosslinked material needed for our proteomic analyses. This brief labeling period did not significantly affect the composition of the transcriptome, as confirmed by RNA sequencing (RNA-seq) (Figure S1A). After 10 min labeling, 4sU was washed away and chased with unmodified uridine until reaching the desired time points. We selected ten time points spanning from 0 to 5 h (0, 15, 30, 45, 60, 90, 120, 180, 240, and 300 min), aiming to cover most stages of the mRNA life cycle based on the reported half-life of human mRNAs (average, 6.9 h; median, 3.4 h).[18,19] Photoactivatable ribonucleoside (PAR)-enhanced crosslinking was induced between 4sU-labeled RNA and its associated proteins in close contact, by exposing the cells to UVA irradiation on ice,[20] and the cells were immediately harvested. After cell lysis, poly(A)+ mRNA molecules were captured with oligo(dT) beads and washed under stringent conditions containing 4 M urea to specifically capture proteins that bind directly to 4sU-labeled mRNAs. For control, we also performed experiments without 4sU labeling (“No-4sU”) to identify non-specific proteins and assess background levels.

To estimate the 4sU incorporation rate under our conditions, we digested RNA into single nucleosides and analyzed them by high-performance liquid chromatography (HPLC) (Figure 1C). Immediately after the 10-min labeling period (chase time = 0 min), 4sU constituted approximately 1.2% of total uridines in the oligo(dT)-captured RNAs. Over the next 30 min, the 4sU fraction increased to 2.3%, likely due to residual intracellular 4sU and delayed phosphorylation of 4sU into 4-thio-uridine triphosphate (4sUTP).[21] Subsequently, this 4sU fraction decreased, possibly as a result of intron removal and decay of labeled transcripts. These results indicate that our pulse-chase conditions labeled approximately 2% of uridine residues, corresponding to around ten 4sU incorporated per mRNA molecule of 2 kb, and the labeling occurred within a relatively narrow time window of about 30–40 min.

Silver staining showed that the captured proteins reached the maximal level at approximately 30 min chase time and then gradually diminished after 60 min (Figure 1D). In the absence of 4sU, protein bands were scarcely visible, which indicates the specificity of 4sU-mediated crosslinking and the stringent washing process used. We conducted experiments in biological triplicate, using approximately 60 million cells for each time point sample. The captured proteins were digested with trypsin, labeled with tandem mass tags (TMT) 11-plex, and analyzed using liquid chromatography tandem mass spectrometry (LC-MS3).[22,23] Protein quantities from triplicate samples were assessed for enrichment against an unlabeled control (Figure S1B).

We identified 801 confidently quantified RBPs by employing two criteria: enrichment over the unlabeled control and reproducibility across replicates (Figure 1E; STAR Methods). To assess the specificity of our method in identifying RBPs, we compared these proteins with previously identified RBPs.[9] More than 90% (734/801) of the confidently quantified proteins in our study matched those previously reported as mRNA interactors, with over 80% of them annotated with “RNA binding” GO term (Figure 1F). Furthermore, we found 58 Pfam domains that were significantly enriched in the confidently quantified proteins over all human proteins, at a false discovery rate (FDR) of 1% (Figure 1G).[24] A majority of these enriched domains (35 out of 58), such as RNA recognition motif (RRM), helicase C, Asp-Glu-Ala-Asp (DEAD), and K homology (KH), were experimentally verified as RNA binding domains (RBDs) or enriched in earlier RIC studies (Table S1).[5,8] Taken together, these results indicate that our method successfully captures mRBPs.

To examine the captured RNAs under our conditions, we performed RNA-seq on oligo(dT)-enriched RNAs at chase times 0, 30, and 300 min (Figures S1C and S1D; see STAR Methods). We also carried out an extra enrichment procedure by using methane thiosulfonate conjugated to biotin (MTS-biotin) to specifically biotinylate and isolate 4sU-labeled RNAs for sequencing. We found no substantial difference between 0 and 30 min. After 300 min, unstable mRNAs (bottom 20% in half-life) showed a modest reduction (from 21.1% to 20.1% in read fractions normalized against the unlabeled spike control) (Figure S1D; see STAR Methods). Nevertheless, the majority (98%) of genes detected at 0 min remained detectable at 300 min (read count per million [RPM] > 10). This indicates that our experimental conditions represent the majority of mRNA species, although we acknowledge that differential mRNA decay rates may influence the mRBP repertoire at later time points.

Overall RNA binding dynamics of RBPs align well with their known functions and localizations

When we examined the confidently quantified and previously reported 734 RBPs, we found substantial variations in their temporal binding patterns. To investigate their RNA binding dynamics quantitatively, we merged normalized protein quantities using a univariate spline (Figure S2A). We defined the “peak binding time” for each RBP as the moment when the spline curve hits its maximum (Figure S2A). This approach facilitated a comparison of the binding dynamics among RBPs. It should be noted that this peak binding time may be different from the true value, owing to the inevitable delay in 4sU uptake, resulting in a lag time of up to 10 min, and also because of 4sU persisting in cells after washing, leading to a residual labeling for about 30 min.

We performed k-means clustering, based on Z score-normalized $\log_2$ protein intensity and peak binding time, resulting in seven temporal clusters (Figure 2A; Table S2; see also STAR Methods). The GO term analyses revealed unique enrichment patterns within each cluster (Figure S2B).[25] For example, cluster I is associated with transcription and 3′ end processing, whereas cluster II shows enrichment in terms related to pre-mRNA splicing. Clusters III and IV are enriched with diverse terms

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

CellPress

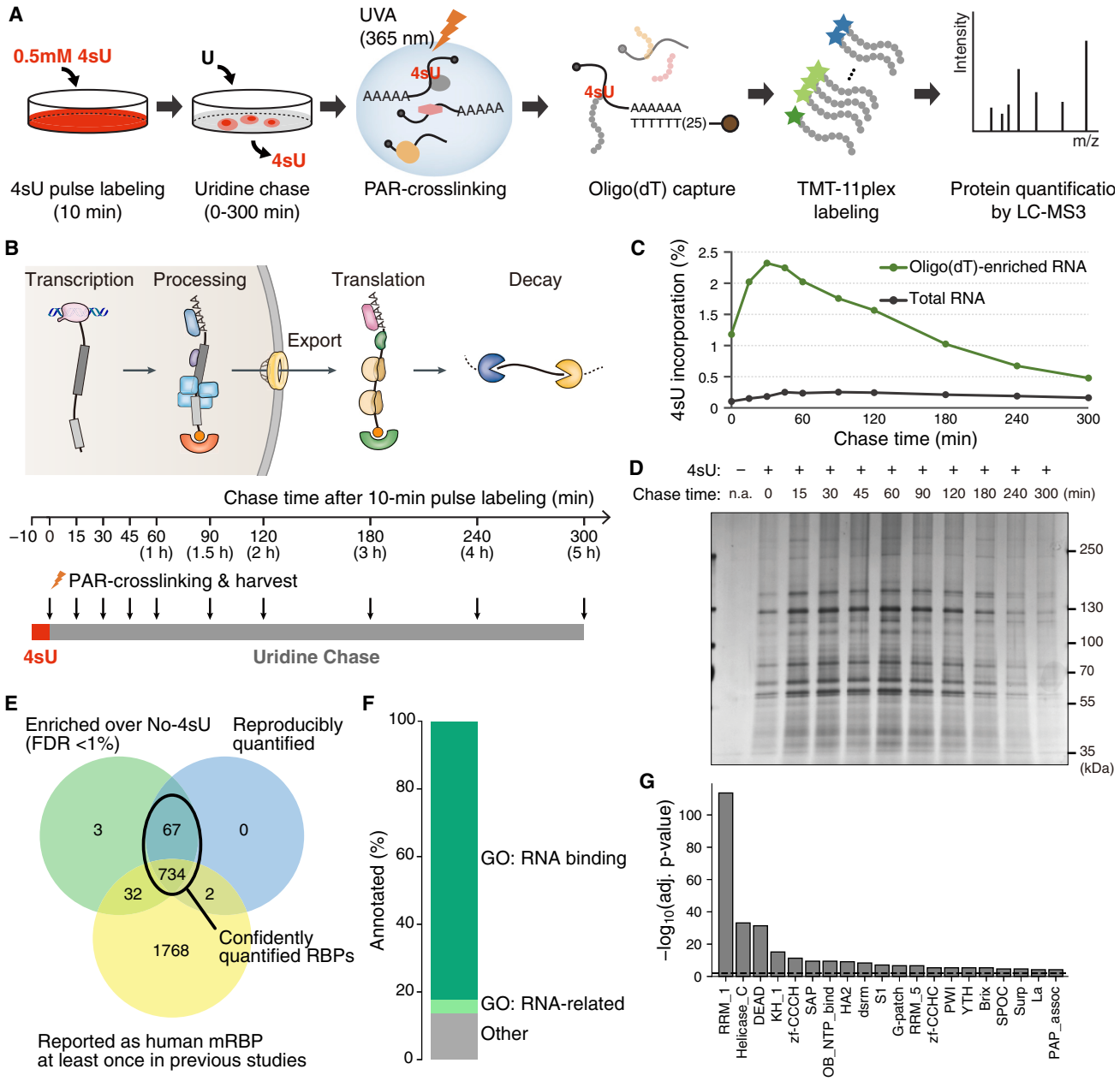


Figure 1. Identification of time-resolved mRNPs using pulse-chase RNA interactome capture (RIC)

(A) Schematic of modified RIC for time-resolved mRNP capture.

(B) Design of high temporal resolution mRNP pulse-chase labeling using 4sU and U nucleosides. Arrows indicate harvest time points.

(C) Proportion of 4sU relative to total U as measured by HPLC after single-nucleotide digestion, using total RNA (black) or oligo(dT)-enriched RNA (green).

(D) Silver staining visualizing the proteins eluted in the RIC experiments.

(E) Venn diagram showing the number of previously reported human mRBPs, proteins enriched over the control (No-4sU) in this study, and of those that are reproducibly quantified.

(F) Gene Ontology (GO) annotations of the confidently quantified RBPs (n = 801). RNA-related GO term was defined as any GO term containing "RNA."

(G) Top 20 enriched Pfam protein domains among the confidently quantified proteins (n = 801). p values were calculated by Fisher's exact test and adjusted by Benjamini-Hochberg method. Dashed line indicates an adjusted p value of 0.01.

including nucleocytoplasmic transport and rRNA processing, and cluster V is linked with nonsense-mediated decay (NMD), mRNA transport, and translation. Finally, clusters VI and VII are marked by translation regulation, stress granule (SG), and mRNA stability terms. Because this initial list included 156 RBPs known to interact with mitochondrial mRNAs, small

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

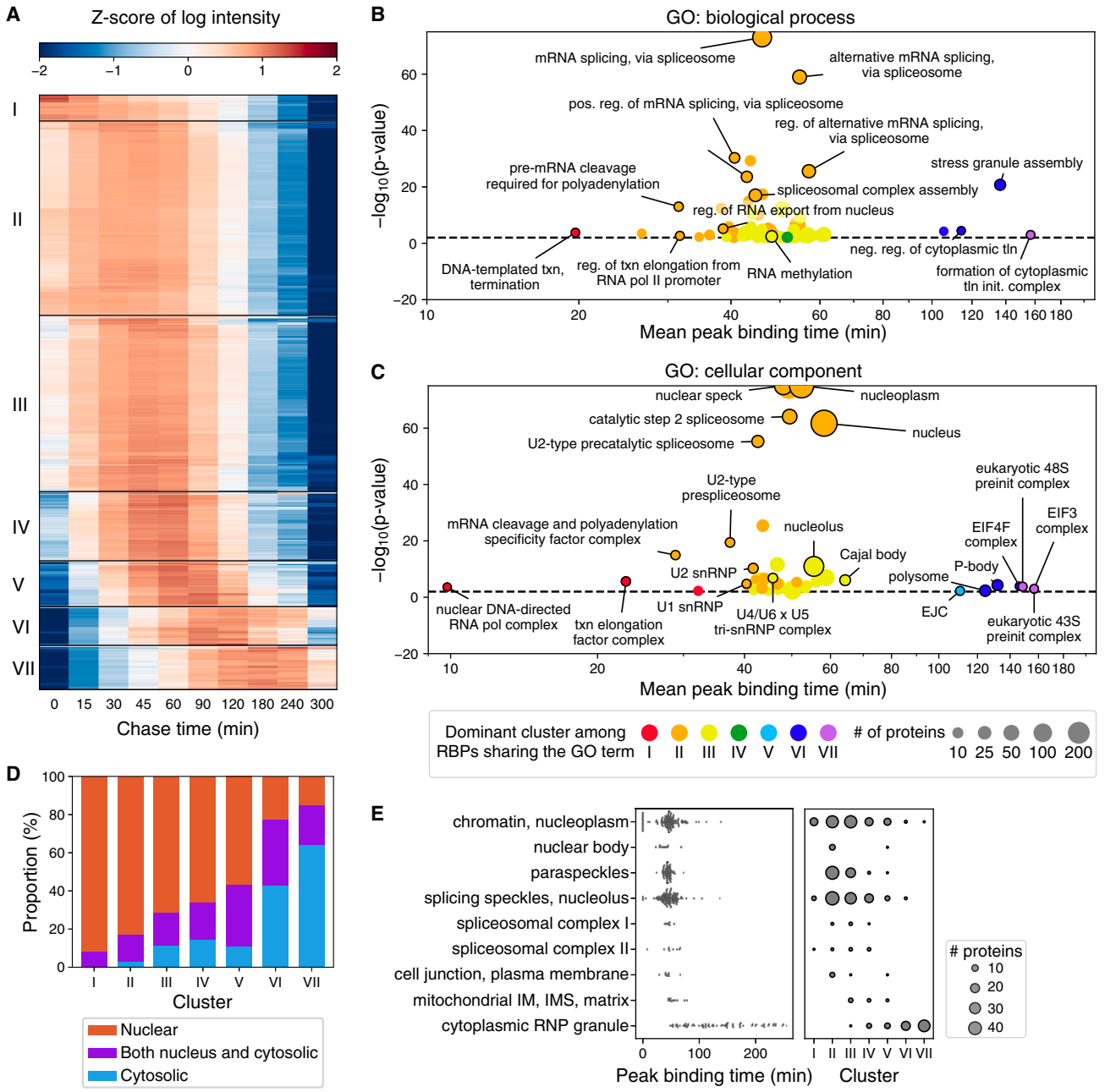


Figure 2. RBP-centric clustering analysis defines temporal stages of mRNP life cycle

(A) Clustering analysis based on mRNA binding dynamics. Each row of the heatmap represents the *Z* score-normalized mRNA binding dynamics of a protein. Out of 801 confidently quantified proteins, 734 proteins that were previously reported as mRNA binders were used for clustering analysis.

(B) Bubble plot showing the mean peak binding time (x axis) of the RBPs sharing the same GO: biological process (BP) terms. The significance of similarity in mRNA binding dynamics of the RBP group sharing the same terms is shown in y axis, as p values derived from two-sided Mann-Whitney U test on the Euclidean distances between proteins with the same GO annotations vs. those between proteins without the same GO annotations. Dashed line indicates p = 0.01. To reduce the over-representation of general (higher level) GO terms, the *elim* algorithm[25] was applied during the p value calculation. Color code indicates the most frequently occurring cluster among the RBPs annotated with each GO term. Area of each circle is proportional to the number of RBPs annotated with each GO. Mitochondrial mRNA binders (n = 65) and potential noncoding RNA binders (n = 65, e.g., ribosomal proteins, snoRNA binders) were excluded from this and following analyses (see STAR Methods for the list).

(C) The same as (B) except for using the GO: cellular component.

(legend continued on next page)

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

nucleolar RNAs (snoRNAs), and rRNAs, which is likely due to polyadenylated mitochondrial mRNAs and abundant noncoding RNAs with internal adenosine-rich sequences (Figures S2C and S2D), we excluded these RBPs from our analysis hereafter, so as to focus on proteins in the nuclear genome-encoded mRNA pathway. However, we note that investigation of the RBPs in the mitochondrial mRNA and noncoding RNA pathways would be interesting topics for future research.

We expanded our analysis on 578 mRBPs by grouping them based on the "biological process (BP)" GO terms and calculating the mean peak binding time of those belonging to the same GO term group. Proteins with transcription-related terms exhibited the earliest mean peak binding times (Figure 2B, shown in x axis) and were primarily affiliated with cluster I or II (Figure 2B, a main cluster indicated by color). Proteins with processing-related terms and translation-related terms also generally followed anticipated temporal orders, albeit with some unexpected observations, such as certain EIF3 subunits (the "formation of cytoplasmic translation initiation complex" GO group) appearing at late time points. The same analysis was applied to "cellular component (CC)" GO terms to observe temporal changes in mRNP organizations (Figure 2C). The mean peak binding times of proteins belonging to RNA polymerase complex and transcription elongation factor complex were the earliest. They were followed by those associated with cleavage and polyadenylation specificity factor (CPSF) complex and spliceosome. Components of exon junction complex (EJC), polysome, P-body (PB), EIF4F complex, and EIF3 complex took longer to bind to mRNA. Thus, the RNA binding dynamics overall align well with known temporal order of mRNP remodeling. Our data suggest that the transition from nuclear mRNP to cytoplasmic mRNP occurs at around 60–120 min chase time. Thus, considering the lag period of approximately 30 min in our labeling scheme, the transition may take 30–90 min after transcription. This is not far from the reported time needed for mRNA export (30–40 min).[26,27]

We further examined subcellular localization of RBPs using the Human Protein Atlas, which is based on immunofluorescence staining experiments.[28] We learned that proteins in cluster I are predominantly nuclear, as expected (Figure 2D). The fraction of nuclear proteins decreases in later clusters while the fractions of cytoplasmic proteins and those that are in both the nucleus and cytoplasm increase. We also cross-referenced proximity labeling-based localization data[29] and found that most early binders are associated with chromatin, nucleoplasm, splicing speckles, and paraspeckles, while late binders are primarily in cytoplasmic RNP granules (Figure 2E).

Target mRNAs of RBPs

Next, we examined RBPs' binding sites on mRNAs. For this, we re-analyzed the enhanced crosslinking and immunoprecipitation (eCLIP) followed by high-throughput sequencing data from the Encyclopedia of DNA Elements (ENCODE) project.[30–32] The eCLIP datasets from K562 and HepG2 cells revealed RNA interaction sites for 85 and 76 confidently quantified RBPs, respectively. We calculated the proportion of eCLIP peaks mapped to the intron, 5′ untranslated region (UTR), coding sequence (CDS), and 3′ UTR (Figure 3A). Although some clusters are represented with only a few eCLIP datasets, we observed consistent patterns from both cell lines. Many early binders in clusters I, II, and III bind mainly to intronic regions, indicating their major roles in pre-mRNPs. In contrast, late binders in clusters V, VI, and VII exhibited fewer eCLIP peaks in introns and a large number of peaks in CDS and 3′ UTR, consistent with their functions in mature mRNA complexes.

When we examined the half-lives of their target mRNAs, we found a weak positive correlation between the RBP's peak binding time and the mean half-life of target mRNAs (Figures 3B and S3A). This suggests that RBPs interacting with stable mRNAs tend to be captured at later time points. LARP4, EIF3G, and SND1 bind to stable mRNAs, suggesting that their temporal binding pattern is mainly determined by their RNA specificity. But some cluster VII proteins bind to a very broad spectrum of targets, including many unstable ones (e.g., DDX3X, UPF1, FXR2, and FAM120A), implying that they may be recruited in a stage-specific manner.

Frequent PPIs between RBPs with similar dynamics

Because RBPs often function within complexes by interacting with other RBPs,[33,34] we analyzed PPIs of all RBPs identified in this study (Biological General Repository for Interaction Datasets [BioGRID][35]). We found that proteins within the same cluster tend to interact frequently with one another (Figure 3C, shown in color). By measuring the number of PPIs between RBPs, we confirmed that the frequency of PPIs within the same cluster or between neighboring clusters were significantly higher than randomly selected protein pairs (Figure 3D). Similar patterns were observed when we included proteins interacting indirectly via one neighbor (Figure S3B). These results suggest that proteins interacting with one another bind to mRNAs at similar time points, via PPIs as well as RNA-protein interactions, collectively forming stage-specific RNP complexes.

To statistically validate this result, we quantified temporal differences in mRNA binding by calculating the Euclidean distance (ED) of RNA binding dynamics between protein pairs (see STAR Methods). Protein pairs with physical interaction evidence show similar RNA binding dynamics, resulting in small EDs, compared with non-interacting protein pairs (Figure S3C). Additionally, we cross-examined known protein complexes (from comprehensive resource of mammalian protein complexes [CORUM], marked as "CORUM complexes")[36] by comparing the distances in mRNA binding dynamics. RBPs within the same complexes (Figure 3E, left, red) displayed significantly smaller EDs, compared with randomly selected proteins (Figure 3E, left, white). We re-analyzed previous data, which classified mRNP complexes

(D) Proportion of nuclear and cytosolic proteins in each cluster, according to the annotation in Human Protein Atlas. "Both" refers to proteins that are known to be located in both the nucleus and the cytosol.

(E) Beeswarm plot (left) and dot plot (right) of subcellular localization of RBPs. The y axis of both plots indicates proximity labeling-based prediction of subcellular location, provided by the HumanCellMap. The x axis of the left plot shows the peak binding time of each protein, and the x axis of the right plot represents the cluster of each circle. The radius of each circle in the right plot is proportional to the number of proteins belonging to each category.

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

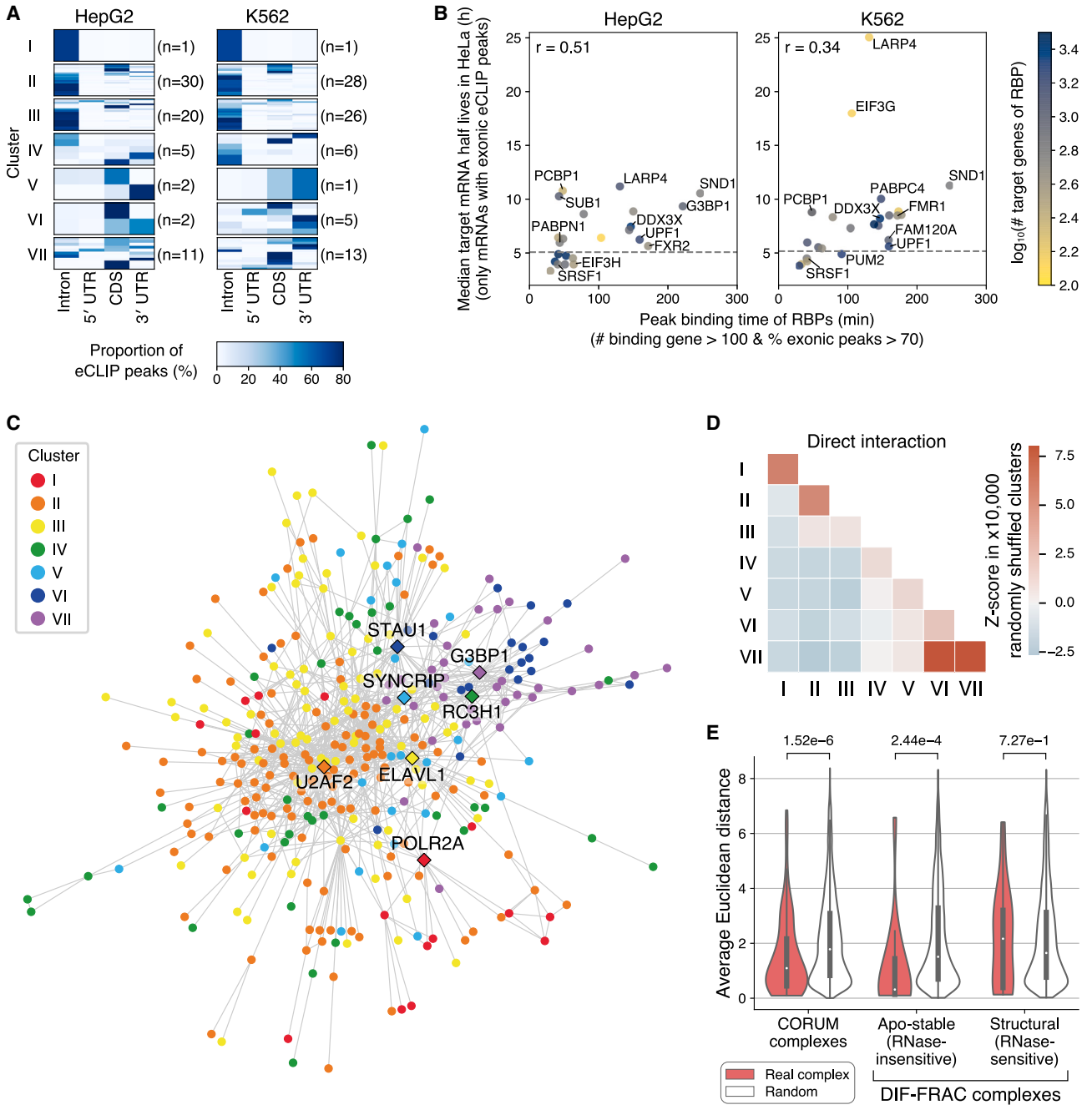


Figure 3. Protein-RNA and protein-protein interactions comply with temporal mRNP remodeling

(A) Location of eCLIP peaks of 76 RBPs in HepG2 (left) and 85 RBPs in K562 (right), which overlap between our data and the ENCODE data. The number of proteins in each cluster is shown on the right.

(B) Scatterplots illustrating the correlation between peak binding time of each RBP (x axis) and median half-lives of eCLIP-identified target mRNAs (y axis, from Tani et al.[18]). Only target mRNAs with exonic eCLIP peaks were considered for median half-life calculation. The color of each dot represents the number of target genes with exonic peaks. Gray dashed line indicates the median half-life of all mRNAs with exonic peaks in any RBP's eCLIP. See Figure S3A for additional information. Pearson's correlation coefficients (r) are indicated in each plot.

(C) A protein-protein interaction (PPI) network of RBPs found in this study (color-coded to indicate their respective clusters). A graph with the largest number of connected proteins (n = 355) was chosen for visualization. The proteins with the most interactions in each cluster are shown as square nodes.

(legend continued on next page)

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

(from differential fractionation [DIF-FRAC] method, marked as "DIF-FRAC" complexes)[33] into two groups based on RNase sensitivity: "apo-stable" RNPs (RNase-insensitive; RNA-independent) and "structural RNPs" (RNase-sensitive; RNA-dependent). The apo-stable RNP components demonstrated significantly smaller distances in RNA dynamics, compared with random protein sets (Figure 3E, middle), while structural RNP components did not (Figure 3E, right). This result implies that proteins interacting with one another independently of RNA tend to join mRNP simultaneously. Overall, PPI and mRNA binding dynamics are in good agreement, indicating that proteins interacting with one another constitute stage-specific RNPs together.

Despite the good overall agreement between PPI and mRNA binding dynamics, some RBPs display markedly different dynamics from their known PPI partners, represented by large EDs (Figure S3D; Table S3). For instance, a late binder MOV10 (cluster VII) has been reported to interact with both early binders (such as XRN2) and late binders (such as UPF1, STAU2, and IGF2BPs) and has been implicated in splicing, translation, and mRNA stability control (Figure S3E). The RNA binding time revealed in this study suggests that the majority of MOV10 molecules act in the late stage of the mRNA life cycle, at least under our experimental conditions.

Early binders involved in transcription and processing

We examined individual representative RBPs with well-established functions (Figure 4A). Large subunits of RNA polymerase II (RNA Pol II) (POLR2A and POLR2B) exhibited the maximal level of interaction at chase time 0 min, then swiftly declined (Figures 4A and 4B). Similar or slightly delayed binding patterns were observed with transcription elongation factor (SUPT5H), anti-termination factors (SCAF4 and SCAF8), histone chaperone facilitates chromatin transcription (FACT) complex components (SSRP1/FACT80 and SUPT16H/FACT140), transcriptional regulators (PHF3 and ZNF579), and the epigenetic regulator human silencing hub (HUSH) complex components (FAM208A/TASOR and PPHLN1). Considering the delayed 4sU incorporation resulting from cellular uptake and the time required for transcriptional elongation (4–12 min for a median-length human gene of 24 kb[37]), it is likely that RBPs enriched at 0 min are associated mostly with newly transcribed transcripts. Given our experimental design using oligo(dT) capture, their presence was unanticipated as these proteins function before the polyadenylation step. To see if the nascent RNAs were captured due to internal A-rich sequences, we analyzed the MTS-biotin-RNA-seq data (Figure S4A). We could not find a significant enrichment of internal A-rich sequences in the labeled RNAs captured at 0 min. Thus, our data are more consistent with a recent proposal that Pol II and newly transcribed transcripts may incorporate into RNP complexes in nuclear matrix, where poly(A)-tailed pre-mRNAs are retained for further processing steps for some period.[38]

At 0 min, we also detected 3′ end processing factors, but their levels increased further to reach their peak binding times at 10–30 min. We detected the components of cleavage stimulation factor (CstF) complex (CSTF3, CSTF2, and CSTF1), CPSF complex (FIP1L1, WDR33, CPSF4, and RBBP6), and cleavage factor Im (CFIm) complex (CPSF6, CPSF7, and CPSF5/NUDT21), which are known to interact with the downstream U/GU-rich element, AAUAAA polyadenylation signal, and upstream U-rich/UGUA-rich element, respectively. It is noted that these factors remain bound to processed mRNAs.[39] In contrast, non-RNA binders and enzymes, such as CPSF1/CPSF160, CPSF2/CPSF100, CPSF3/CPSF73, Symplekin, poly(A) polymerase, and the CFIIm components, were not detected in our experiment, consistent with the notion that these factors do not directly or stably interact with mRNA (Figure S4B).[39]

Compared with the 3′ end processing factors, splicing-related factors showed slightly delayed dynamics (Figures 4C and S4C, peak binding times 20–40 min). Given that our method enriches polyadenylated RNAs, our result indicates that a proportion of spliceosomes may assemble posttranscriptionally after polyadenylation. We detected the U1 and U2 small nuclear ribonucleoprotein particle (snRNP) components and the A complex proteins slightly earlier than the U4/U6.U5 tri-snRNP components and proteins associated with the B, B_{act}, and C complexes (Figure S4C). The peak binding times of U1 and U2 components are 7 min earlier on average than those of tri-snRNP components ($p = 0.002$, two-sided t test). This result is consistent with the known order of spliceosome assembly.

Nuclear pre-mRNA/mRNA binding proteins, such as nuclear cap binding protein 2 (NCBP2/CBP20), hnRNPs, SR proteins, and nuclear PABP (PABPN1), peaked at similar or slightly later time points (approximately 45 min), compared with splicing factors (Figure 4C). Immediately following these proteins, we found the nuclear export factor 1 (NXF1) (peak binding time, 50 min).

Intriguingly, the transcription-export (TREX) complex components (ALYREF/THOC4, UAP56/DDX39B, CHTOP, and SARNP/CIP29) known to recruit NXF1 were not detected until later (maximal peak times, 67–77 min) (Figure 4D), suggesting that the majority of NXF1 molecules may be recruited independently of TREX. At comparable time points, many other proteins reported to be involved in mRNA export are also recruited: RBM33, ZC3H11A, FYTTD1/UIF, POLDIP3/SKAR, NCBP3, API5, CCDC9, and LRPPRC (Figure S4D). Our result indicates that the export-competent mRNP is assembled mainly at 60–70 min chase time (considering the labeling time, this may correspond to 30–80 min posttranscription). Thus, a drastic reconfiguration may take place at this point to generate an mRNP with distinct composition and structure.

The high temporal resolution of our data allowed us to observe differences between related proteins. For instance, while most hnRNPs, which bind and regulate pre-mRNAs, showed similar dynamics (Figure S4E), one notable exception was HNRPQ

(D) PPIs within and between clusters. To calculate the normalized enrichment of PPIs, *Z* scores were calculated through 10,000 iterations of randomly shuffling clusters. This generated null distributions for PPI counts, from which the *Z* scores were derived.

(E) Violin plots of average Euclidean distances of mRNA binding dynamics, per protein complex. p values were calculated by the two-sided Mann-Whitney U test. Boxplots inside the violin plots show the median (center dot), first and third quartiles (lower and upper box limits, respectively), and 1.5 times the interquartile range (whiskers). CORUM: all protein complexes in the CORUM database. DIF-FRAC complexes: RNA binding protein complexes, defined by Mallam et al.[33]

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

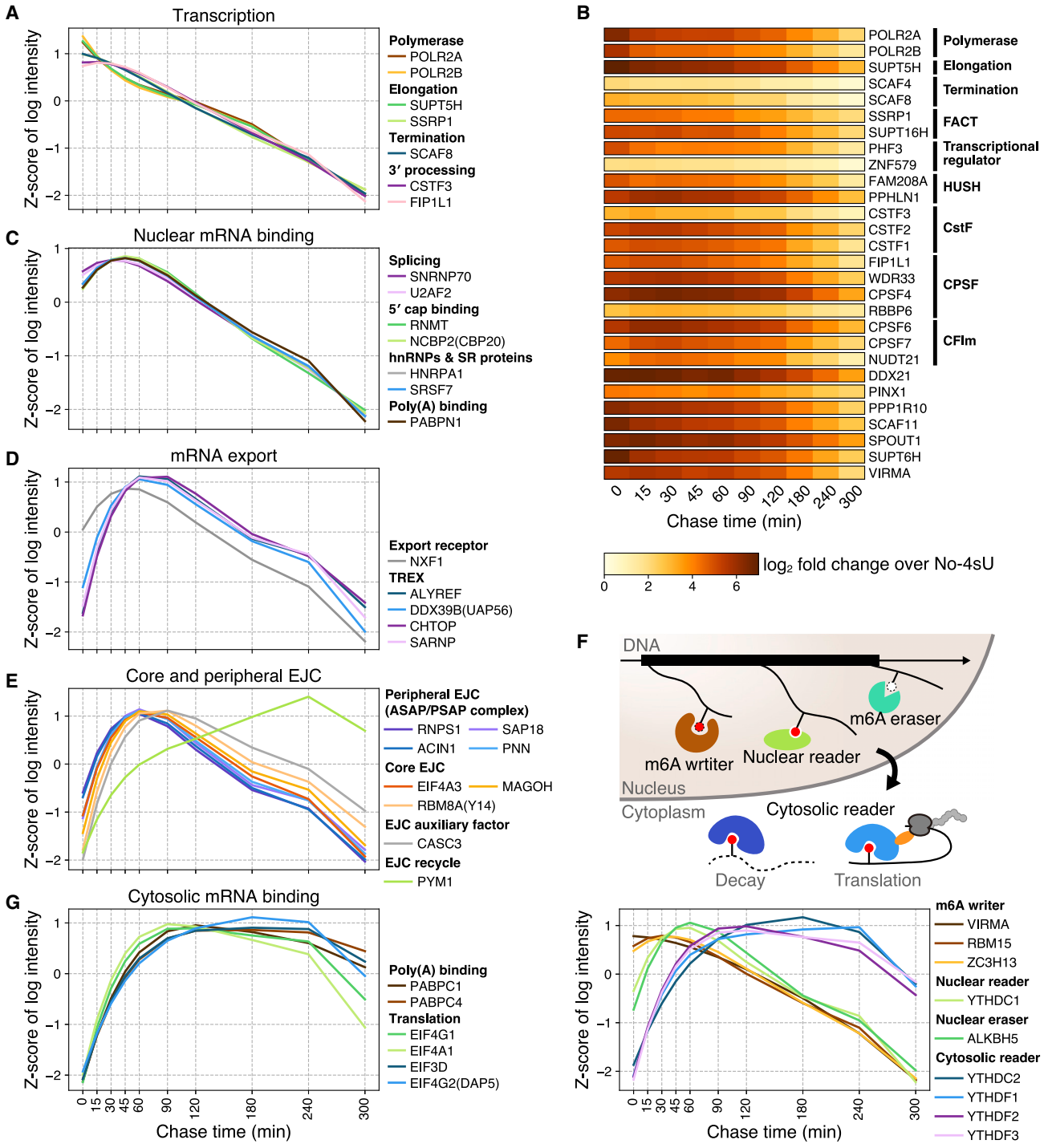


Figure 4. Expected and unexpected mRNA binding dynamics of mRBPs

(A) mRNA binding dynamics of RBPs known to participate in mRNA transcription.

(B) mRNA binding dynamics of RBPs known to participate in transcription or putative mRNA transcription regulators enriched at 0 min (peak binding time = 0), shown as $\log_2$ fold change over control (No-4sU).

(C–E) mRNA binding dynamics of RBPs known to bind nuclear mRNAs (C), participate in mRNA export (D), and compose exon junction complexes (EJCs, E).

(F) Schematic diagram of m6A RNA modification and mRNA binding dynamics of m6A RNA modification-related proteins.

(G) mRNA binding dynamics of RBPs known to bind cytosolic mRNAs.

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

(also known as hnRNP Q or SYNCRIP), detected substantially later than the other hnRNPs. This observation supports previous findings that HNRPQ/SYNCRIP modulates mRNA translation and decay.[40,41]

Furthermore, the core EJC components (EIF4A3, MAGOH, and RBM8A/Y14) precede CASC3/MLN51/BTZ that is not essential for EJC formation and mainly located in the cytoplasm. PYM1 that functions in EJC recycling in the cytoplasm is indeed detected substantially later than the core EJC subunits (Figure 4E).[42]

Factors involved in m6A modification present another interesting example of dynamic mRNA interaction (Figure 4F). The nuclear writer complex components (VIRMA, RBM15, and ZC3H13) are followed by the nuclear reader (YTHDC1), nuclear eraser (ALKBH5), and cytoplasmic readers (e.g., YTHDF2, YTHDF3, YTHDC2, and YTHDF1).[43,44] The writer complex components were detected earlier than EJC (Figures 4E and 4F), suggesting that the loading of the m6A writer may begin before splicing and EJC loading. Whether or not the modification reaction takes place before EJC assembly is currently unclear and will be an interesting topic to investigate.

mRNP remodeling in the cytoplasm

Cytoplasmic PABPs (PABPC1 and PABPC4) increase gradually, reaching the maximal level at 120–140 min (Figure 4G). In part, this result may reflect large variations in the rates of export and cytosolic remodeling between mRNA species. Alternatively, but not mutually exclusively, PABPC molecules may accumulate on poly(A) tail incrementally rather than all at once after nuclear export. Note that the fourth RRM motif of PABPCs have low specificity, allowing their binding and crosslinking to the 3′ UTR.[45] High-density loading of PABPCs on poly(A) tail would allow frequent interaction between the 3′ UTR and PABPC.

Translation factors like EIF4A1 and EIF4G peak at around 100 min chase time (corresponding to 70–110 min posttranscription), indicating active translation by this point (Figure 4G). EIF4E, the cytosolic canonical cap binding protein, was not detected in our experiment, likely due to its poor crosslink ability. We identified EIF3D, alongside its co-factor EIF4G2/DAP5, exhibiting slightly delayed dynamics, compared with EIF4G/EIF4G1, which implies their preferential binding to aged mRNAs and/or their specificity to a subset of mRNAs.

Posttranscriptional suppressors such as the Argonaute proteins (AGO2 and AGO3), Pumilio proteins (PUM1 and PUM2), and YTHDF proteins (YTHDF2 and YTHDF3) also emerged at similar time points (Figure S4F), suggesting that they may start to impact translation and deadenylation soon after nuclear export.

We noticed that the majority of RBPs show a single peak, but some proteins bind to RNA across a broad range of chase time rather than at a specific time point (Figure S4G). GEMIN5 is known to function in snRNP assembly through its N-terminal domain and regulates mRNA translation via its C-terminal domain.[46] GEMIN5 seems to peak both at 30–45 and 240 min, reflecting its multiple roles. Similarly, SND1 and NFX1, also known to be multifunctional, exhibited broad binding patterns.

Our current findings help us rethink the current model of mRNP remodeling (Figure 5A). For validation of mRNA binding dynamics, we conducted western blot analyses after 4sU-labeling, PAR-crosslinking, and oligo(dT) capture (Figure 5B). Consistent with the mass spectrometry data, the Pol II subunit POLR2 was detected primarily at the earliest time point, which was followed by nuclear RBPs such as hnRNPA1, SRSF7, PABPN, and TDP-43. The export receptor NXF1 appeared prior to eIF4A3, the EJC subunit, and ALYREF, the TREX component. Cytosolic translation factors, PABPC4 and EIF4G2, were detected later. Late binders such as LARP1, FMRP, and G3BP1 reached the highest levels at 4–5 h. Thus, the results from western blotting confirmed our mass spectrometry-based dataset.

Aged mRNPs are enriched with SG proteins

It was intriguing to us that the clusters VI and VII are highly enriched with GO terms related to "cytoplasmic RNP granules" and "SG assembly" (Figures 2B and 2E), even though we did not expose the cells to any stressors. When we performed immunofluorescence experiments with an antibody against G3BP1, an SG marker, we did not see any discernible foci (Figure 6A), indicating that SGs were not formed under our experimental conditions.

Despite the absence of SGs, we found that a majority (63%) of cluster VII proteins overlap with "G3BP1 interactome"[47] (Figure 6B), which include G3BP1, FMR1, LSM14A, and UBAP2/UBAP2L. Most of these proteins have been reported to bind to G3BP1 independently of stress.[47] We further compared our RBPs with those known to localize to cytosolic granules such as SGs and PBs.[48–50] Remarkably, 67% and 74% of cluster VI and VII proteins were described as SG proteins, respectively, while PB proteins are not strongly enriched in late clusters (Figure 6C; Table S4). Furthermore, 26% of cluster VII proteins were annotated as the regulators of SG formation as well as the structural core proteins of SG (Figure 6D). Out of 36 proteins previously reported as SG regulators and core proteins, 16 proteins belong to cluster VI or VII (Figure 6E). Therefore, aged mRNPs detected in our experiments are similar to SGs in its protein composition.

We observed frequent PPIs within and between clusters VI and VII (Figures 3D and 6F). The majority of these interacting proteins are indeed SG proteins, while those not participating in the interaction networks are predominantly non-SG proteins (Figures 6F and 6G). The considerable overlap between the SG proteome and late mRNPs, along with the frequent PPIs among late binders, suggests that late-stage mRBPs may form submicroscopic RNPs with similar properties to SGs.

Interaction between viral RNAs and late binders

Cluster VII also contains many RBPs that are known to interact with RNA viruses. Proteomics-based approaches on RNA viruses have identified a large number of RBPs known to interact with viral RNAs (viral RBPs [vRBPs]).[51–55] We found that 354 RBPs identified in our study have previously been described as vRBPs (Figure S5A; Table S4).

All seven temporal clusters contain vRBPs; however, we observed marked enrichment of vRBPs among late binders, most notably in cluster VII (Figure S5B). Transcripts of single-stranded positive-sense RNA viruses such as coronaviruses (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

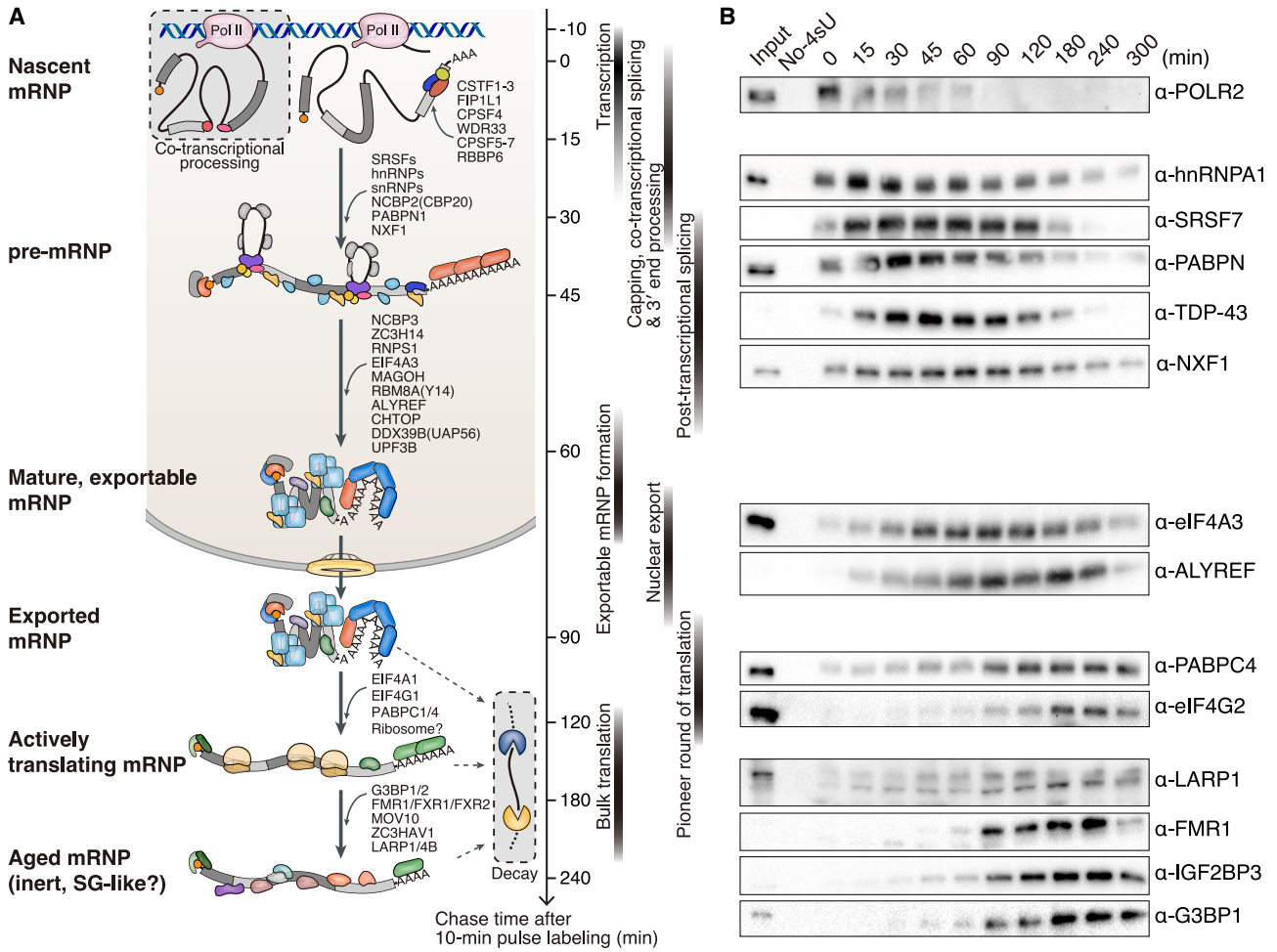


Figure 5. A model for mRNP remodeling through mRNA life cycle and validation of mRNA binding dynamics with western blots

(A) Suggested model of mRNP remodeling through mRNA life cycle, based on mRNA binding dynamics.

(B) Validation by western blotting. Input represents 0.01% of the lysate that is used in the 0 min RIC sample. No-4sU means the RIC elution sample obtained from unlabeled cells. All blots were generated with the same set of samples except hnRNPA1, which was loaded at one-third the amount of the other samples.

and human coronavirus OC43 [HCoV-OC43]) and flaviviruses (dengue virus and zika virus), which replicate within the cytoplasm, showed particularly strong enrichment with late clusters. This enrichment is not solely attributed to the cytosolic localization of late binders, as the same analysis restricted to the cytosolic proteins also exhibited higher enrichment in clusters VI and VII (Figure S5C). Moreover, we noticed that SG proteins frequently appeared among vRBPs in clusters VI and VII (Figure S5D). The SG protein proportion among vRBPs was higher than that among the host mRNA interactome across all clusters. In the intersection among cluster VII, SG, and vRBP, we found well-known antiviral proteins such as DDX3X and ZC3HAV1/ZAP. These associations imply the roles of the granule-forming late binders in viral life cycle.

Systemic identification of RBPs with unexpected RNA binding dynamics

While mRNA binding dynamics align well with previously reported characteristics in general, some RBPs showed significant discrepancies between their RNA binding times and known functions, localizations, and PPIs, implying undiscovered functions for these RBPs. To systematically identify RBPs with unexpected dynamics, we developed a regression method that predicts mRNA binding dynamics based on annotated characteristics (Figure 7A; for detailed information, see the STAR Methods section). We compiled a gene-GO term table that includes the RBPs identified in this study along with their corresponding GO term annotations. Due to the redundancy of certain GO terms (e.g., RNA binding and "nucleic acid binding"), we applied multiple correspondence analysis (MCA) to compress the information into a lower dimension. Subsequently, with the MCA-converted GO annotations, we fitted a ridge regression model to predict the Z score of a given RBP's quantity at each time point.

The coefficient of determination (R^2) between the observed and expected Z scores at each time point ranged 0.55–0.79 (Figure S6A). Despite the overall high performance of our prediction model on well-annotated RBPs (Figure S6B; Table S5), we

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

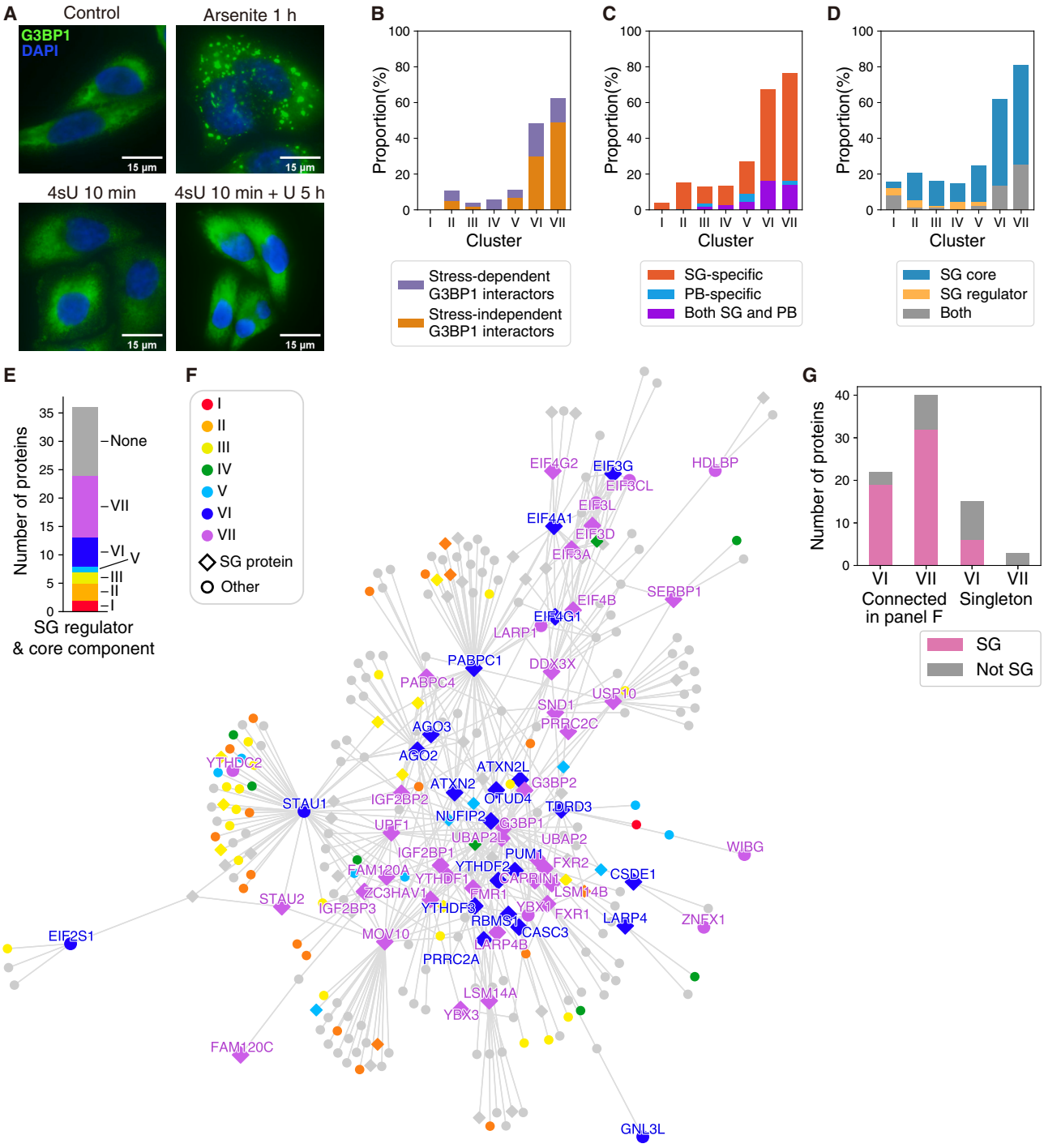


Figure 6. Stress granule proteins are enriched in aged mRNPs

(A) Immunofluorescence against G3BP1 (green) and DAPI (blue) in HeLa cells. Sodium arsenite treatment was used as a positive control for stress granule (SG) formation.

(B) Proportion of stress-dependent (purple) or stress-independent (orange) G3BP1 interactors in each cluster. The list of G3BP1 protein interactors were from Markmiller et al.[47]

(C) Proportion of SG and P-body (PB) proteins in each cluster. The tier 1 SG and PB lists from the RNA granule database (v1.0) were utilized.

(D) Proportion of SG core constituents (enriched in SG pull-down proteome) and SG regulators (necessary for SG formation, identified by genetic screening) that belong to each cluster. "Both" indicates proteins that were reported as both SG core and SG regulators. The protein lists were obtained from Yang et al.[48]

(legend continued on next page)

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

observed some RBPs with substantial differences between the expected and observed mRNA binding dynamics (Figures 7B–7D and S6C; Table S5). Under-characterized RBPs, sparsely covered in literature, generally displayed high prediction errors (Figure S6D). We sorted RBPs based on the difference between the expected and observed dynamics (Figures 7B–7D). RBPs at the top or bottom of this list bind to mRNA earlier or later than our model predicts. For example, LSM14B, FAM120A, and FAM120C, which are poorly studied and associated with only a few GO terms (''ribonucleoprotein complex,'' ''mRNA binding,'' RNA binding, and ''regulation of translation''), were predicted as an intermediate binder (cluster V), but they actually associate with mRNAs at very late time points (cluster VII). The dynamics of these RBPs suggests their function in the last stage of mRNA life cycle. Thus, our temporal RNA interaction data not only supplement current knowledge but are also useful for identifying under-characterized RBPs.

DISCUSSION

This study is the first longitudinal proteomic analysis of mRNPs with a high temporal resolution. Although numerous studies have been performed for decades to assign the functions and localizations of RBPs, from which the sequential orders of mRNA binding have been inferred, direct and quantitative data have not been available. The high resolution of our data allows us to detect even subtle differences in mRNA binding dynamics, establishing the chronological orders of RNA-protein interactions throughout the mRNA life cycle.

To foster community access to our data, we developed an interactive web application available at https://chronology.rna.snu.ac.kr (Figure S7A). This platform enables the search for specific RBPs via their UniProt accessions or gene symbols. For each RBP, we provide detailed and comprehensive information on mRNA binding dynamics, subcellular localization, GO-based prediction of mRNA binding dynamics, and protein interactors. Furthermore, our web application allows researchers to compare the mRNA binding dynamics of multiple RBPs of interest and to create combined line plots of the mRNA binding dynamics. This tool is very useful for comparing proteins with related functions. In addition, our web application may assist the studies on RBPs implicated in human diseases (Figure S7B). For instance, TDP43, EWS, TAF15, FUS, and hnRNPA1 whose mutations are implicated in amyotrophic lateral sclerosis belong to cluster II or III, indicating their original roles in nuclear pre-mRNP processing. FMRP, responsible for fragile X syndrome, belongs to cluster VII, along with its paralogs FXR1 and FXR2, which is consistent with their proposed role as translational repressors.[56] An oncogene, SND1, has been implicated in multiple steps from transcription and splicing to mRNA stabilization. SND1 showed a broad binding pattern in the late time points after 90 min, indicating its major role as a posttranscriptional regulator.

The temporal sequence of RBPs' binding is largely consistent with their subcellular location and molecular interaction data. However, there were many unanticipated findings. Particularly interesting was the strong overlap between cluster VII and SG proteome, which includes FMR1/FXR1/FXR2, G3BP1/2, UPF1, MOV10, DDX3X, YBX1/3, LSM14A/B, STAU2, PURA/B, IGF2BP1/2/3, UBAP2/UBAP2L, FAM120A/C, LARP1/4B, SND1, and ZC3HAV1. Recent studies have shown that SG components interact with one another and co-localize in the cytoplasm even in the absence of stress, implying that there might be a ''pre-SG'' complex in normal cell condition.[47,57] One can envision that at least part of these proteins are recruited to old mRNAs, forming submicroscopic complexes. This RNP assembly might be part of the natural process of mRNA aging even under unstressed conditions. Given that some of these proteins are implicated in translational repression and decay, this aged mRNP complex may be in a ''retired'' state, in which mRNA translation is less active, compared with those in ''younger'' mRNPs. Because SG formation per se does not prevent translation,[58] the assembly of the SG proteins on old mRNAs is likely to be a consequence, rather than a cause, of the natural process of ''translational retirement.'' Whether or not this complex is indeed less active in translation, what triggers the changes in mRNP composition, how ''old'' mRNA is distinguished from ''young'' mRNAs, and if liquid-liquid phase separation is involved in this process will be interesting topics for future studies.

It is also noteworthy that this very late-binding group is highly enriched with proteins that bind to viral transcripts (vRBPs), particularly those from coronaviruses and flaviviruses, which possess mRNA-like positive-sense single-stranded RNA genomes. Such vRBPs, including DDX3X and ZC3HAV1/ZAP, may play roles in antiviral defense, possibly by sequestering viral RNAs, while some of them could be repurposed by viruses to facilitate viral proliferation.[59,60] It will be interesting to investigate the functions of these late-binding RBPs in the viral infection cycle.

Another intriguing observation involves nuclear export factors. The export receptor NXF1 is known to be recruited by ''adapter proteins'' like TREX and SR proteins (SRSF3/SRp20 and SRSF7/9G8).[61,62] Our current data show that NXF1 is categorized under cluster III and is recruited shortly after SR proteins (mainly cluster II) but before TREX (cluster V) (Figures 4D and 4E). NXF1 can directly interact with mRNAs without adapter molecules *in vitro*, and it has been detected in many mRNA interactome capture studies.[9] Ultraviolet light C (UVC)-based CLIP sequencing (CLIP-seq) analyses revealed that NXF1 binds broadly to exons, with a modest enrichment in the 3′ UTR near SRSF binding sites.[63] This implies that SRSF proteins might assist NXF1's direct binding to RNA. It is worth noting that UVC-based CLIP-seq data show that the splicing index of target RNAs of NXF1 is comparable to that of the SRSF proteins (Figures S8A and S8B),[63] while ALYREF shows a stronger preference for spliced mRNAs (exon-exon boundary) compared

(E) Number of SG proteins that belong to different temporal clusters. SG regulator and SG core protein were from Yang et al.[48]
(F) Interaction network map of proteins belonging to cluster VI, cluster VII, and their interactors (n = 298). SG proteins (SG tier 1 in RNA granule database [v1.0]) are marked as square nodes. Gene names of cluster VI and VII proteins are marked above their respective nodes.
(G) The number of SG and non-SG proteins among the connected cluster VI, VII proteins in (F) and not connected (singleton) cluster VI, VII proteins.

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

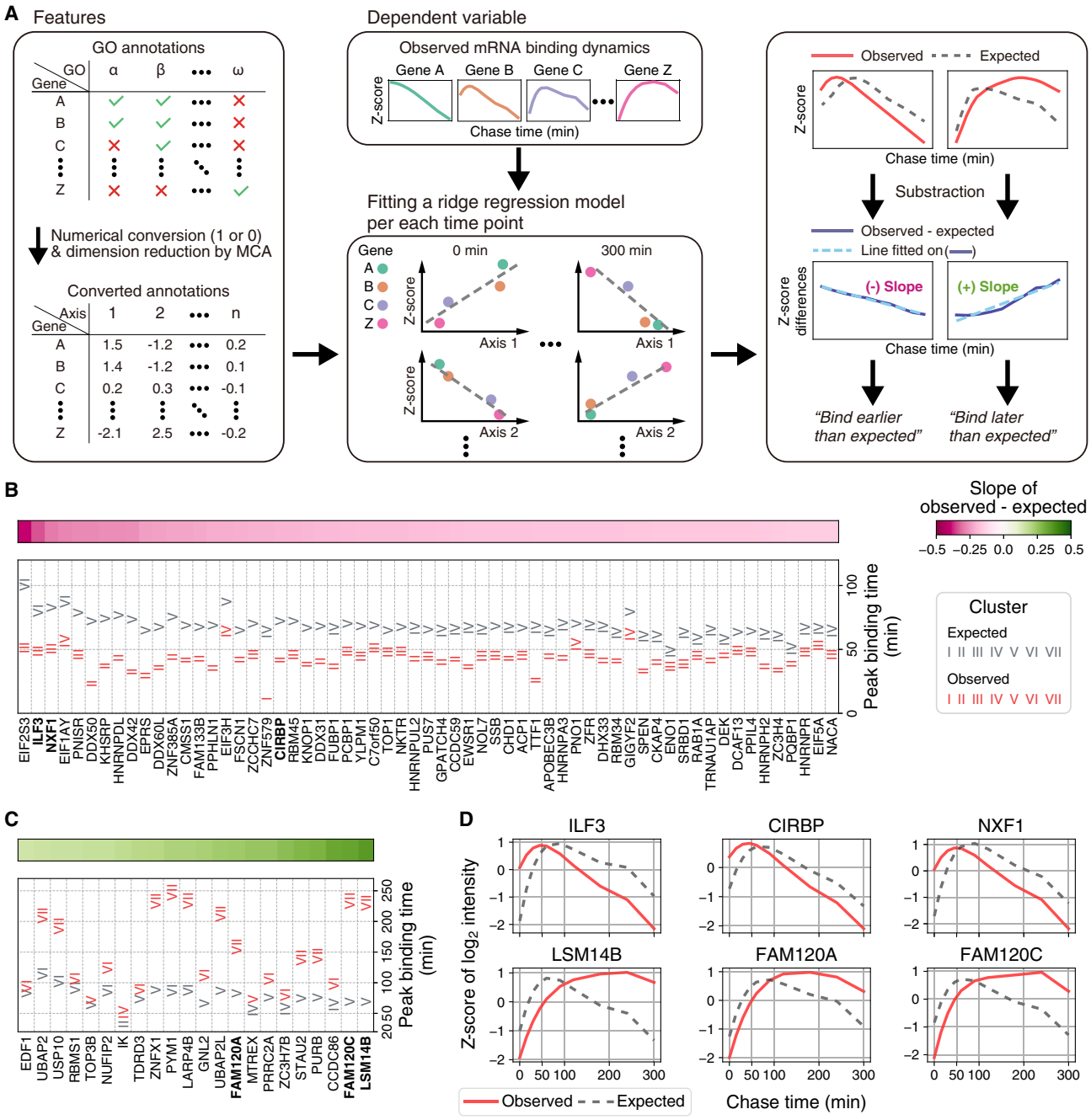


Figure 7. RBPs of unexpected mRNA binding dynamics

(A) Schematic of the mRNA binding dynamics prediction model based on GO-term annotations. The gene-GO term table is encoded to numbers and transformed to the lower dimensions, by the multiple correspondence analysis (MCA, left column). For each time point, a ridge regression model is fitted to find the relationship between the MCA-converted GO annotations and *Z* score-normalized intensities. Model fitting was repeated for all 10 time points (middle column). The difference between the observed and expected values reveals RBPs that bind earlier or later than predicted (right column).

(B and C) Unexpected early (B) or late (C) binders. Shown are RBPs whose slope of the observed-minus-expected *Z* score is higher than 0.15 or lower than −0.15.

(D) Observed and expected mRNA binding dynamics of selected RBPs.

with NXF1 (Figures S8C and S8D).[63–65] Taken together, NXF1 might be largely recruited before ALYREF. While we do not exclude the potential role of TREX in NXF1 recruitment, it seems that TREX primarily functions either at the final stages of nuclear mRNP remodeling or binds to a subset of mRNAs, which are processed more slowly and extensively (Figure 5A).[66]

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Importantly, the EJC core proteins (EIF4AIII, MAGOH, and RBM8A/Y14) and the TREX components (ALYREF/THOC4, UAP56/DDX39B, CHTOP, and SARNP/CIP29) show distinctly delayed dynamics (as cluster V, at 60–70 min chase time), compared with 3′ processing factors (clusters I or II), PABPN (cluster II), spliceosome components (mainly cluster II), and SR proteins (cluster II). The components of apoptosis- and splicing-associated protein (ASAP) complex (RNPS1, SAP18, and ACIN1) and its alternate component PNN, which are known as peripheral EJC proteins, belong to cluster IV and bind to mRNA slightly earlier than EJC (Figure 4E). Because transcription-related factors are strongly reduced by 30–45 min and PABPN reaches its maximal level at 30–45 min, it is highly likely that ASAP, EJC, and TREX join the mRNP primarily after the completion of splicing and polyadenylation. These findings are surprising given that TREX is known to function in conjunction with RNA polymerase and is recruited co-transcriptionally.[67] Our data demonstrate that human mRNP undergoes a drastic posttranscriptional remodeling after the completion of transcription and RNA processing. A large number of RBPs, including ASAP, EJC, and TREX, join the processed mRNP at 60–70 min chase time (which corresponds to about 30–80 min after transcription). This remodeling step might allow the compaction and preparation of mRNP for nuclear export and be largely "uncoupled" from transcription. This finding challenges the current view of mechanistic "coupling" between transcription and export.

Other cluster IV or V proteins may also play important roles in mRNP remodeling and export. A notable example is ZC3H14, a deeply conserved nuclear PABP whose homolog, Nab2, is known to be required for poly(A) tail length control and mRNA export in yeast, fly, and human.[68] Although Nab2 was thought to couple multiple processes from transcription, polyadenylation, and export, a recent in-depth analysis showed that human Nab2/ZC3H14 interacts with EJC and TREX and that it mediates mRNA export,[69] which is in line with our temporal data. Two nuclear PABPs, PABPN1 (cluster II) and ZC3H14 (cluster IV), show different dynamics, indicating distinct roles. It awaits further studies as to if and how these two proteins co-occupy the same mRNA and which role each protein plays (Figures 4C and S4D). Apart from ZC3H14, clusters IV and V contain many other proteins implicated in mRNA export, including RBM33 (IV), ZC3H11A (IV), FYTTD1/UIF (V), POLDIP3/SKAR (V), NCBP3 (V), API5 (V), CCDC9 (V), and LRPPRC (V).[70–73] These proteins may constitute and mark the export-competent mRNP, which is likely to be protein-rich and compact. By comparing spliceosome B_{act} complex and TREX, we can estimate that it takes approximately 25 min to form the export-competent mRNP complex after posttranscriptional splicing (Figure 5A). Our time-resolved data highlight the need to re-visit the mechanisms of nuclear mRNP remodeling and mRNA export. It will be interesting to find out how these RBPs are organized and interact with one another to construct a compact structure. Our discovery of RBPs marking this stage will help purify and examine the composition, structure, and function of the mature export-competent mRNP.

In this study, we found many RBPs with unanticipated binding dynamics. To systematically identify them, we trained a machine learning model to predict mRNA binding dynamics from GO annotations. This model was then applied to screen under-characterized RBPs by comparing the observed and expected dynamics. We found numerous RBPs that bind to mRNA earlier or later than expected, revealing proteins with potentially unknown functions and/or multiple functions. For instance, we discovered ILF3/NF90 as an early binder in cluster III (Figure 7D). ILF3 has been implicated in various processes such as transcription, microRNA maturation, pre-mRNA splicing, RNA export, translation, and mRNA degradation.[74,75] However, recent eCLIP experiments showed a high proportion of intron peaks (Figure S7C), suggesting ILF3's binding to pre-mRNAs rather than mature mRNAs. This aligns with our data and suggests a role of ILF3 mainly in pre-mRNA processing. Another example is CIRBP identified as an early binder in cluster II (Figure 7D). Although CIRBP has been described as a regulator of mRNA stabilization and translational activation,[76,77] it is predominantly in the nucleus and partially relocates to the cytoplasm only under stress conditions.[78] Our result supports its nuclear function in unstressed conditions. Our annotation-based prediction method and the temporal RNA binding data will provide a useful resource for RBP research.

The methodology used in this study is readily adaptable to other biological contexts, provided that pulse chasing and UVA crosslinking are feasible. For instance, a loss-of-function study could be combined with this method to explore an RBP's function in mRNP remodeling. Given that some RBPs are associated with genetic disorders, examining the impact of their mutations in disease models could be informative. One could also scrutinize the effects of pharmacological inhibitors/agonists on mRNP remodeling. Investigating the molecular effects of splicing or translation inhibitors, currently under clinical trials or in use for cancer and genetic diseases, could be enlightening.[79–82] Moreover, temporal mRNP profiling during cell reprogramming and under stress conditions could reveal potential alterations in mRNP remodeling in cells coping with differentiation signals and stressors, such as amino acid deprivation, endoplasmic reticulum (ER) stress, heat shock, viral infection, and inflammation. This approach could introduce a time dimension to the understanding of posttranscriptional gene regulation.

We have excluded RBPs with well-established functions in the mitochondrial RNA and rRNA pathways from our analyses. However, these proteins were quantitated reproducibly in our experiments, providing useful additional information. Proteins involved in mitochondrial mRNA transcription and processing factors bind to RNA earlier than decay factors and stability regulators, as expected (Figure S7D).[83] We could also quantitate RNA binding dynamics of nucleolar RBPs, which are overall consistent with their proposed sub-organellar localizations (Figure S7E).[84] Given their abundance, time-resolved profiling is readily attainable for mitochondrial RNAs, rRNAs, and small nuclear RNAs (snRNAs) and therefore could offer valuable insights into the complex processes of mitochondrial gene regulation, ribosome biogenesis, and spliceosome biogenesis.

Limitations of the study

As we used PAR-crosslinking, our data do not fully capture RBPs that bind to RNA in uridine-depleted regions or those that cannot be crosslinked efficiently by UVA. In the future, orthogonal

crosslinking methods will help us to gain a more comprehensive view of RNP remodeling.

The current study utilized an oligo(dT) capture method to analyze poly(A)+ mRNAs. Thus, our protocol depletes mRNAs that either lack a poly(A) tail (such as replication-dependent histone mRNAs) or possess very short poly(A) tails (such as nascent transcripts and decay intermediates). As a result, the RBPs quantified in this study do not fully represent co-transcriptional processing factors or decay factors (Figure 5A, indicated within dotted boxes). Future studies might consider using organic phase separation-based methods to capture total RNPs, although this could introduce its own limitations due to a lack of RNA specificity.[85,86]

Ideally, future studies should aim to isolate and analyze gene-specific mRNPs, using specific antisense oligos as baits, to overcome the limitations of bulk analysis. Bulk mRNA capture used in this study does not account for gene-specific regulatory mechanisms. Also, with bulk analyses, the proportion of stable mRNAs modestly increased over time, enriching RBPs that preferentially associate with stable mRNAs at later time points. Currently, gene-specific capture presents a major technical challenge due to the low copy numbers of individual mRNA species and the detection limits of mass spectrometry. Therefore, a substantial technical improvement in mass spectrometry is needed to overcome these limitations.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.03.012.

ACKNOWLEDGMENTS

We thank members of our laboratory for discussion and technical support, especially Hyerim Yi, Myeonghwan Kim, Ahyeon Son, and Sanghee Shin for material and protocols. We are also grateful to Hyeshik Chang for the support on the computational resources. This work was supported by grants IBS-R008-D1 from the Institute for Basic Science from the Ministry of Science, ICT and Future Planning of Korea (Y.C., B.U., Y.N., J.K., J.-S.K., and V.N.K.) and NRF-2022R1C1C2003822 from the National Research Foundation of Korea (Y.N.) and by BK21 Research Fellowships from the Ministry of Education of Korea (B.U.).

AUTHOR CONTRIBUTIONS

V.N.K. conceived the project and collected financial support. B.U., Y.C., Y.N., and J.-S.K. designed the mRNP capture and mass spectrometry experiments. B.U., J.K., and J.-S.K. generated the LC-MS3 data, and Y.N. performed HPLC experiments. Y.C. conducted bioinformatic analyses. Y.C., B.U., and V.N.K. wrote the manuscript.

DECLARATION OF INTERESTS

V.N.K. is a member of the *Molecular Cell* advisory board.

Received: January 23, 2024
Revised: February 16, 2024
Accepted: March 14, 2024
Published: April 8, 2024

REFERENCES

1. Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem. *84*, 325–354. https://doi.org/10.1146/annurev-biochem-080111-092106.
2. Mitchell, S.F., and Parker, R. (2014). Principles and properties of eukaryotic mRNPs. Mol. Cell *54*, 547–558. https://doi.org/10.1016/j.molcel.2014.04.033.
3. Gall, J.G. (1956). Small granules in the amphibian oocyte nucleus and their relationship to RNA. J. Biophys. Biochem. Cytol. *2* (*Suppl*), 393–396. https://doi.org/10.1083/jcb.2.4.393.
4. Dreyfuss, G. (1986). Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. Annu. Rev. Cell Biol. *2*, 459–498. https://doi.org/10.1146/annurev.cb.02.110186.002331.
5. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian

mRNA-binding proteins. Cell *149*, 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031.

6. Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690. https://doi.org/10.1016/j.molcel.2012.05.021.

7. Na, Y., Kim, H., Choi, Y., Shin, S., Jung, J.H., Kwon, S.C., Kim, V.N., and Kim, J.S. (2021). FAX-RIC enables robust profiling of dynamic RNP complex formation in multicellular organisms in vivo. Nucleic Acids Res. *49*, e28. https://doi.org/10.1093/nar/gkaa1194.

8. Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130. https://doi.org/10.1038/nsmb.2638.

9. Caudron-Herger, M., Jansen, R.E., Wassmer, E., and Diederichs, S. (2021). RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. Nucleic Acids Res. *49*, D425–D436. https://doi.org/10.1093/nar/gkaa1040.

10. Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845. https://doi.org/10.1038/nrg3813.

11. Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A., and Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat. Commun. *7*, 12128. https://doi.org/10.1038/ncomms12128.

12. Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A.I., González-Almela, E., Lenz, C.E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T., et al. (2019). System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. Mol. Cell *74*, 196–211.e11. https://doi.org/10.1016/j.molcel.2019.01.017.

13. Perez-Perri, J.I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., and Hentze, M.W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nat. Commun. *9*, 4408. https://doi.org/10.1038/s41467-018-06557-8.

14. Zhang, Z., Liu, T., Dong, H., Li, J., Sun, H., Qian, X., and Qin, W. (2021). An RNA tagging approach for system-wide RNA-binding proteome profiling and dynamics investigation upon transcription inhibition. Nucleic Acids Res. *49*, e65. https://doi.org/10.1093/nar/gkab156.

15. Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat. Methods *15*, 213–220. https://doi.org/10.1038/nmeth.4595.

16. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141. https://doi.org/10.1016/j.cell.2010.03.009.

17. Kleiner, R.E. (2021). Interrogating the transcriptome with metabolically incorporated ribonucleosides. Mol. Omics *17*, 833–841. https://doi.org/10.1039/d1mo00334h.

18. Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. *22*, 947–956. https://doi.org/10.1101/gr.130559.111.

19. Singer, R.H., and Penman, S. (1973). Messenger RNA in HeLa cells: kinetics of formation and decay. J. Mol. Biol. *78*, 321–334. https://doi.org/10.1016/0022-2836(73)90119-8.

20. Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-linking of nucleic acids to associated proteins. Crit. Rev. Biochem. Mol. Biol. *32*, 101–140. https://doi.org/10.3109/10409239709108550.

21. Duffy, E.E., Schofield, J.A., and Simon, M.D. (2019). Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. Wiley Interdiscip. Rev. RNA *10*, e1513. https://doi.org/10.1002/wrna.1513.

22. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. *75*, 1895–1904. https://doi.org/10.1021/ac0262560.

23. Ting, L., Rad, R., Gygi, S.P., and Haas, W. (2011). MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat. Methods *8*, 937–940. https://doi.org/10.1038/nmeth.1714.

24. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230. https://doi.org/10.1093/nar/gkt1223.

25. Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607. https://doi.org/10.1093/bioinformatics/btl140.

26. Jelinek, W., Adesnik, M., Salditt, M., Sheiness, D., Wall, R., Molloy, G., Philipson, L., and Darnell, J.E. (1973). Further evidence on the nuclear origin and transfer to the cytoplasm of polyadenylic acid sequences in mammalian cell RNA. J. Mol. Biol. *75*, 515–532. https://doi.org/10.1016/0022-2836(73)90458-0.

27. Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat. Cell Biol. *12*, 543–552. https://doi.org/10.1038/ncb2056.

28. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321. https://doi.org/10.1126/science.aal3321.

29. Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., et al. (2021). A proximity-dependent biotinylation map of a human cell. Nature *595*, 120–124. https://doi.org/10.1038/s41586-021-03592-2.

30. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711–719. https://doi.org/10.1038/s41586-020-2077-3.

31. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514. https://doi.org/10.1038/nmeth.3810.

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

33. Mallam, A.L., Sae-Lee, W., Schaub, J.M., Tu, F., Battenhouse, A., Jang, Y.J., Kim, J., Wallingford, J.B., Finkelstein, I.J., Marcotte, E.M., et al. (2019). Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. Cell Rep. *29*, 1351–1368.e5. https://doi.org/10.1016/j.celrep.2019.09.060.

34. Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., Kettenbach, A.N., and Diederichs, S. (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. Mol. Cell *75*, 184–199.e10. https://doi.org/10.1016/j.molcel.2019.04.018.

35. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535–D539. https://doi.org/10.1093/nar/gkj109.

36. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. Nucleic Acids Res. *47*, D559–D563. https://doi.org/10.1093/nar/gky973.

37. Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. *15*, R69. https://doi.org/10.1186/gb-2014-15-5-r69.

38. Tang, P., Yang, Y., Li, G., Huang, L., Wen, M., Ruan, W., Guo, X., Zhang, C., Zuo, X., Luo, D., et al. (2022). Alternative polyadenylation by sequential activation of distal and proximal PolyA sites. Nat. Struct. Mol. Biol. *29*, 21–31. https://doi.org/10.1038/s41594-021-00709-z.

39. Boreikaitė, V., and Passmore, L.A. (2023). 3′-End Processing of Eukaryotic mRNA: Machinery, Regulation, and Impact on Gene Expression. Annu. Rev. Biochem. *92*, 199–225. https://doi.org/10.1146/annurev-biochem-052521-012445.

40. Svitkin, Y.V., Yanagiya, A., Karetnikov, A.E., Alain, T., Fabian, M.R., Khoutorsky, A., Perreault, S., Topisirovic, I., and Sonenberg, N. (2013). Control of translation and miRNA-dependent repression by a novel poly(A) binding protein, hnRNP-Q. PLoS Biol. *11*, e1001564. https://doi.org/10.1371/journal.pbio.1001564.

41. Weidensdorfer, D., Stöhr, N., Baude, A., Lederer, M., Köhn, M., Schierhorn, A., Buchmeier, S., Wahle, E., and Hüttelmaier, S. (2009). Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. RNA *15*, 104–115. https://doi.org/10.1261/rna.1175909.

42. Schlautmann, L.P., and Gehring, N.H. (2020). A Day in the Life of the Exon Junction Complex. Biomolecules *10*, 866. https://doi.org/10.3390/biom10060866.

43. Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. Nat. Rev. Genet. *15*, 293–306. https://doi.org/10.1038/nrg3724.

44. Jiang, X., Liu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., Yang, C., and Chen, Y. (2021). The role of m6A modification in the biological functions and diseases. Signal Transduct. Target. Ther. *6*, 74. https://doi.org/10.1038/s41392-020-00450-x.

45. Webster, M.W., Chen, Y.H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Coller, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. Mol. Cell *70*, 1089–1100.e8. https://doi.org/10.1016/j.molcel.2018.05.033.

46. Martinez-Salas, E., Embarc-Buh, A., and Francisco-Velilla, R. (2020). Emerging Roles of Gemin5: From snRNPs Assembly to Translation Control. Int. J. Mol. Sci. *21*, 3868. https://doi.org/10.3390/ijms21113868.

47. Markmiller, S., Soltanieh, S., Server, K.L., Mak, R., Jin, W., Fang, M.Y., Luo, E.C., Krach, F., Yang, D., Sen, A., et al. (2018). Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. Cell *172*, 590–604.e13. https://doi.org/10.1016/j.cell.2017.12.032.

48. Yang, P., Mathieu, C., Kolaitis, R.M., Zhang, P., Messing, J., Yurtsever, U., Yang, Z., Wu, J., Li, Y., Pan, Q., et al. (2020). G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. Cell *181*, 325–345.e28. https://doi.org/10.1016/j.cell.2020.03.046.

49. Youn, J.Y., Dyakov, B.J.A., Zhang, J., Knight, J.D.R., Vernon, R.M., Forman-Kay, J.D., and Gingras, A.C. (2019). Properties of Stress Granule and P-Body Proteomes. Mol. Cell *76*, 286–294. https://doi.org/10.1016/j.molcel.2019.09.014.

50. Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487–498. https://doi.org/10.1016/j.cell.2015.12.038.

51. Lee, S., Lee, Y.S., Choi, Y., Son, A., Park, Y., Lee, K.M., Kim, J., Kim, J.S., and Kim, V.N. (2021). The SARS-CoV-2 RNA interactome. Mol. Cell *81*, 2838–2850.e6. https://doi.org/10.1016/j.molcel.2021.04.022.

52. Ooi, Y.S., Majzoub, K., Flynn, R.A., Mata, M.A., Diep, J., Li, J.K., van Buuren, N., Rumachik, N., Johnson, A.G., Puschnik, A.S., et al. (2019). An RNA-centric dissection of host complexes controlling flavivirus infection. Nat. Microbiol. *4*, 2369–2382. https://doi.org/10.1038/s41564-019-0518-2.

53. Knoener, R., Evans, E., 3rd, Becker, J.T., Scalf, M., Benner, B., Sherer, N.M., and Smith, L.M. (2021). Identification of host proteins differentially associated with HIV-1 RNA splice variants. eLife *10*, e62470. https://doi.org/10.7554/eLife.62470.

54. Kim, B., Arcos, S., Rothamel, K., Jian, J., Rose, K.L., McDonald, W.H., Bian, Y., Reasoner, S., Barrows, N.J., Bradrick, S., et al. (2020). Discovery of Widespread Host Protein Interactions with the Pre-replicated Genome of CHIKV Using VIR-CLASP. Mol. Cell *78*, 624–640.e7. https://doi.org/10.1016/j.molcel.2020.04.013.

55. Iselin, L., Palmalux, N., Kamel, W., Simmonds, P., Mohammed, S., and Castello, A. (2022). Uncovering viral RNA-host cell interactions on a proteome-wide scale. Trends Biochem. Sci. *47*, 23–38. https://doi.org/10.1016/j.tibs.2021.08.002.

56. Darnell, J.C., and Klann, E. (2013). The translation of translational control by FMRP: therapeutic targets for FXS. Nat. Neurosci. *16*, 1530–1536. https://doi.org/10.1038/nn.3379.

57. Cho, N.H., Cheveralls, K.C., Brunner, A.D., Kim, K., Michaelis, A.C., Raghavan, P., Kobayashi, H., Savy, L., Li, J.Y., Canaj, H., et al. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. Science *375*, eabi6983. https://doi.org/10.1126/science.abi6983.

58. Mateju, D., Eichenberger, B., Voigt, F., Eglinger, J., Roth, G., and Chao, J.A. (2020). Single-Molecule Imaging Reveals Translation of mRNAs Localized to Stress Granules. Cell *183*, 1801–1812.e13. https://doi.org/10.1016/j.cell.2020.11.010.

59. Onomoto, K., Yoneyama, M., Fung, G., Kato, H., and Fujita, T. (2014). Antiviral innate immunity and stress granule responses. Trends Immunol. *35*, 420–428. https://doi.org/10.1016/j.it.2014.07.006.

60. Bidet, K., Dadlani, D., and Garcia-Blanco, M.A. (2014). G3BP1, G3BP2 and CAPRIN1 are required for translation of interferon stimulated mRNAs and are targeted by a dengue virus non-coding RNA. PLoS Pathog. *10*, e1004242. https://doi.org/10.1371/journal.ppat.1004242.

61. Huang, Y., Gattoni, R., Stévenin, J., and Steitz, J.A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Mol. Cell *11*, 837–843. https://doi.org/10.1016/S1097-2765(03)00089-3.

62. Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Séraphin, B., Wilm, M., Bork, P., and Izaurralde, E. (2000). REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. RNA *6*, 638–650. https://doi.org/10.1017/s1355838200000078.

63. Müller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K., and Neugebauer, K.M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. Genes Dev. *30*, 553–566. https://doi.org/10.1101/gad.276477.115.

64. Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. Mol. Cell *75*, 310–323.e8. https://doi.org/10.1016/j.molcel.2019.04.034.

65. Cordiner, R.A., Dou, Y., Thomsen, R., Bugai, A., Granneman, S., and Heick Jensen, T. (2023). Temporal-iCLIP captures co-transcriptional RNA-protein interactions. Nat. Commun. *14*, 696. https://doi.org/10.1038/s41467-023-36345-y.

66. Zuckerman, B., Ron, M., Mikl, M., Segal, E., and Ulitsky, I. (2020). Gene Architecture and Sequence Composition Underpin Selective

Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex. Mol. Cell *79*, 251–267.e6. https://doi.org/10.1016/j.molcel.2020.05.013.

67. Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Söding, J., Holstege, F.C., and Sträßer, K. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. PLoS Genet. *9*, e1003914. https://doi.org/10.1371/journal.pgen.1003914.
68. Kelly, S.M., Leung, S.W., Pak, C., Banerjee, A., Moberg, K.H., and Corbett, A.H. (2014). A conserved role for the zinc finger polyadenosine RNA binding protein, ZC3H14, in control of poly(A) tail length. RNA *20*, 681–688. https://doi.org/10.1261/rna.043984.113.
69. Morris, K.J., and Corbett, A.H. (2018). The polyadenosine RNA-binding protein ZC3H14 interacts with the THO complex and coordinately regulates the processing of neuronal transcripts. Nucleic Acids Res. *46*, 6561–6575. https://doi.org/10.1093/nar/gky446.
70. Bong, S.M., Bae, S.H., Song, B., Gwak, H., Yang, S.W., Kim, S., Nam, S., Rajalingam, K., Oh, S.J., Kim, T.W., et al. (2020). Regulation of mRNA export through API5 and nuclear FGF2 interaction. Nucleic Acids Res. *48*, 6340–6352. https://doi.org/10.1093/nar/gkaa335.
71. Dou, Y., Barbosa, I., Jiang, H., Iasillo, C., Molloy, K.R., Schulze, W.M., Cusack, S., Schmid, M., Le Hir, H., LaCava, J., et al. (2020). NCBP3 positively impacts mRNA biogenesis. Nucleic Acids Res. *48*, 10413–10427. https://doi.org/10.1093/nar/gkaa744.
72. Thomas, A., Rehfeld, F., Zhang, H., Chang, T.C., Goodarzi, M., Gillet, F., and Mendell, J.T. (2022). RBM33 directs the nuclear export of transcripts containing GC-rich elements. Genes Dev. *36*, 550–565. https://doi.org/10.1101/gad.349456.122.
73. Drew, K., Wallingford, J.B., and Marcotte, E.M. (2021). hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multiprotein assemblies. Mol. Syst. Biol. *17*, e10016. https://doi.org/10.15252/msb.202010016.
74. Castella, S., Bernard, R., Corno, M., Fradin, A., and Larcher, J.C. (2015). Ilf3 and NF90 functions in RNA biology. Wiley Interdiscip. Rev. RNA *6*, 243–256. https://doi.org/10.1002/wrna.1270.
75. Grasso, G., Higuchi, T., Mac, V., Barbier, J., Helsmoortel, M., Lorenzi, C., Sanchez, G., Bello, M., Ritchie, W., Sakamoto, S., et al. (2020). NF90 modulates processing of a subset of human pri-miRNAs. Nucleic Acids Res. *48*, 6874–6888. https://doi.org/10.1093/nar/gkaa386.
76. Yang, C., and Carrier, F. (2001). The UV-inducible RNA-binding protein A18 (A18 hnRNP) plays a protective role in the genotoxic stress response. J. Biol. Chem. *276*, 47277–47284. https://doi.org/10.1074/jbc.M105396200.
77. Wellmann, S., Bührer, C., Moderegger, E., Zelmer, A., Kirschner, R., Koehne, P., Fujita, J., and Seeger, K. (2004). Oxygen-regulated expression of the RNA-binding proteins RBM3 and CIRP by a HIF-1-independent mechanism. J. Cell Sci. *117*, 1785–1794. https://doi.org/10.1242/jcs.01026.
78. De Leeuw, F., Zhang, T., Wauquier, C., Huez, G., Kruys, V., and Gueydan, C. (2007). The cold-inducible RNA-binding protein migrates from the nucleus to cytoplasmic stress granules by a methylation-dependent mechanism and acts as a translational repressor. Exp. Cell Res. *313*, 4130–4144. https://doi.org/10.1016/j.yexcr.2007.09.017.
79. Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. Genes Dev. *17*, 419–437. https://doi.org/10.1101/gad.1048803.
80. Silvera, D., Formenti, S.C., and Schneider, R.J. (2010). Translational control in cancer. Nat. Rev. Cancer *10*, 254–266. https://doi.org/10.1038/nrc2824.
81. Lee, S.C., and Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. Nat. Med. *22*, 976–986. https://doi.org/10.1038/nm.4165.
82. Bosco, D.A. (2018). Translation dysregulation in neurodegenerative disorders. Proc. Natl. Acad. Sci. USA *115*, 12842–12844. https://doi.org/10.1073/pnas.1818493115.
83. Nicholls, T.J., Rorbach, J., and Minczuk, M. (2013). Mitochondria: mitochondrial RNA metabolism and human disease. Int. J. Biochem. Cell Biol. *45*, 845–849. https://doi.org/10.1016/j.biocel.2013.01.005.
84. Shan, L., Xu, G., Yao, R.W., Luan, P.F., Huang, Y., Zhang, P.H., Pan, Y.H., Zhang, L., Gao, X., Li, Y., et al. (2023). Nucleolar URB1 ensures 3′ ETS rRNA removal to prevent exosome surveillance. Nature *615*, 526–534. https://doi.org/10.1038/s41586-023-05767-5.
85. Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat. Biotechnol. *37*, 169–178. https://doi.org/10.1038/s41587-018-0001-2.
86. Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391–403.e19. https://doi.org/10.1016/j.cell.2018.11.004.
87. Adusumilli, R., and Mallick, P. (2017). Data Conversion with ProteoWizard msConvert. Methods Mol. Biol. *1550*, 339–368. https://doi.org/10.1007/978-1-4939-6747-6_23.
88. Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277. https://doi.org/10.1038/ncomms6277.
89. Zhu, Y., Orre, L.M., Zhou Tran, Y., Mermelekas, G., Johansson, H.J., Malyutina, A., Anders, S., and Lehtiö, J. (2020). DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis. Mol. Cell. Proteomics *19*, 1047–1057. https://doi.org/10.1074/mcp.TIR119.001646.
90. Poole, W., Gibbs, D.L., Shmulevich, I., Bernard, B., and Knijnenburg, T.A. (2016). Combining dependent P-values with an empirical adaptation of Brown's method. Bioinformatics *32*, i430–i436. https://doi.org/10.1093/bioinformatics/btw438.
91. Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972. https://doi.org/10.1261/rna.1136108.
92. Garibaldi, A., Carranza, F., and Hertel, K.J. (2017). Isolation of Newly Transcribed RNA Using the Metabolic Label 4-Thiouridine. Methods Mol. Biol. *1648*, 169–176. https://doi.org/10.1007/978-1-4939-7204-3_13.
93. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic acids research *47*, D442–D450. https://doi.org/10.1093/nar/gky1106.
94. Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods *14*, 1198–1204. https://doi.org/10.1038/nmeth.4435.
95. Zhang, B., Chambers, M.C., and Tabb, D.L. (2007). Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. J. Proteome Res. *6*, 3549–3557. https://doi.org/10.1021/pr070230d.
96. Duffy, E.E., and Simon, M.D. (2016). Enriching s^4 U-RNA Using Methane Thiosulfonate (MTS) Chemistry. Curr. Protoc. Chem. Biol. *8*, 234–250. https://doi.org/10.1002/cpch.12.
97. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

98. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

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

100. 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.

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-LARP1	Bethyl	Cat# A302-087A; RRID:AB_1604274
Anti-ALYREF	Bethyl	Cat# A302-892A; RRID:AB_10663772
Anti-TDP-43	Proteintech	Cat# 10782-2-AP; RRID:AB_615042
Anti-G3BP1	BD Biosciences	Cat# 611126; RRID:AB_398437
Anti-SRSF7	MBL International	Cat# RN079PW; RRID:AB_11161213
Anti-FMR1	MBL International	Cat# RN016P; RRID:AB_1953044
Anti-PABPN	Abcam	Cat# ab75855; RRID:AB_1310538
Anti-PABPC4	Thermo Fisher Scientific	Cat# A301-466A; RRID:AB_999661
Anti-POLR2	Santa Cruz Biotechnology	Cat# sc-56767; RRID:AB_785522
Anti-eIF4G2	MBL International	Cat# RN003P; RRID:AB_1570636
Anti-IGF2BP3	Santa Cruz Biotechnology	Cat# sc-365640; RRID:AB_10847223
Anti-NXF1	Abcam	Cat# ab129160; RRID:AB_11142853
Anti-EIF4A3	Proteintech	Cat# 17504-1-AP; RRID:AB_2097393
Alexa Fluor 488 Donkey anti-Mouse IgG (H+L)	Thermo Scientific	Cat# A-21202; RRID:AB_141607
Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson ImmunoResearch Laboratories	115-035-146; RRID:AB_2307392
Peroxidase AffiniPure Goat Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch Laboratories	111-035-144; RRID:AB_2307391
Anti-ALYREF	Bethyl	Cat# A302-892A; RRID:AB_10663772
Anti-TDP-43	Proteintech	Cat# 10782-2-AP; RRID:AB_615042
Anti-G3BP1	BD Biosciences	Cat# 611126; RRID:AB_398437
Anti-SRSF7	MBL International	Cat# RN079PW; RRID:AB_11161213
Anti-FMR1	MBL International	Cat# RN016P; RRID:AB_1953044
Anti-PABPN	Abcam	Cat# ab75855; RRID:AB_1310538
Anti-PABPC4	Thermo Fisher Scientific	Cat# A301-466A; RRID:AB_999661
Anti-POLR2	Santa Cruz Biotechnology	Cat# sc-56767; RRID:AB_785522
Anti-eIF4G2	MBL International	Cat# RN003P; RRID:AB_1570636
Anti-IGF2BP3	Santa Cruz Biotechnology	Cat# sc-365640; RRID:AB_10847223
Anti-NXF1	Abcam	Cat# ab129160; RRID:AB_11142853
Anti-EIF4A3	Proteintech	Cat# 17504-1-AP; RRID:AB_2097393
Chemicals, peptides, and recombinant proteins		
FBS	Welgene	Cat#S001-01
DMEM, High glucose	Welgene	Cat#LM001-05
4-thiouridine	Sigma-Aldrich	Cat#T4509
Uridine	Sigma-Aldrich	Cat#U3003
TURBO DNase I	Invitrogen	Cat#AM2239
SUPERase•In RNase Inhibitor	Invitrogen	Cat#AM2696
UltraPure 1 M Tris-HCl Buffer, pH 7.5	Thermo Scientific	Cat#15567027
Lithium chloride 8 M solution	Sigma-Aldrich	Cat#L7026
Lithium dodecyle sulfate	Sigma-Aldrich	Cat#L9781
Nonidet P 40 Substitute	Sigma-Aldrich	Cat#74385
Urea	Sigma-Aldrich	Cat#U6504
DTT	Sigma-Aldrich	Cat#43819
EDTA (0.5 M), pH 8.0, RNase-free	Invitrogen	Cat#AM9261
Oligo(dT) bead	NEB	Cat#S1419S

(*Continued on next page*)

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HEPES	Sigma-Aldrich	Cat#H0887
Iodoacetamide	Sigma-Aldrich	Cat#I1149
Ammonium bicarbonate	Sigma-Aldrich	Cat#09830
TRIzol	Invitrogen	Cat#15596-018
RNase A	Thermo Scientific	Cat#EN0531
Benzonase Nuclease	Sigma-Aldrich	Cat#E1014
Bacterial Alkaline Phosphatase	Invitrogen	Cat#18011015
Phosphodiesterase I	Worthington Biochem	Cat#LS003926
Novex WedgeWell 10%Tris-Glycine Mini Gels	Invitrogen	Cat#XP00102BOX
SuperSignal West Pico PLUS Chemiluminescent	Thermo Scientific	Cat#34580
SuperSignal West Pico femto Chemiluminescen	Thermo Scientific	Cat#34905
WSE-7240 EzReprobe	ATTO JAPAN	Cat#2332530
8-chamber slide glass	Thermo Scientific	Cat#154461PK
Pierce 16% Formaldehyde (w/v), Methanol-free	Thermo Scientific	Cat#28908
Triton® X-100	Promega	Cat#H5141
DAPI	Sigma-Aldrich	Cat#D9542
MTSEA biotin-XX	Biotium	Cat#90066
N,N-Dimethylformamide	Sigma-Aldrich	Cat#227056
RNAclean XP	Beckman	Cat#A63987
E. coli poly (A) polymerase	NEB	Cat#M0276
MGIEasy RNA Directional Library Prep Kit V2.0	MGI	Cat#1000006385
Critical commercial assays		
Ezway Silver staining kit	Komabiotech	Cat#K14040D
BCA Protein Assay Reagent kit	Pierce	Cat#23227
Deposited data		
Raw image data	This paper	https://doi.org/10.5281/zenodo.10688611
Raw and analyzed proteomic data	This paper	PRIDE: PXD039054
Raw and analyzed RNA-seq data	This paper	GEO: GSE256124
Protein-protein interaction data	Stark et al.[35]	BioGRID release 4.2.191
Subcellular localization data	Go et al.[29]	Human cell map database v1
Subcellular localization data	Thul et al.[28]	Human Protein Atlas, downloaded on March 8, 2019
Protein domain database	Finn et al.[24]	Pfam, version 32.0
RNA interactome database	Caudron-Herger et al.[9]	RBP2GO database
eCLIP	ENCODE consortium	https://www.encodeproject.org/
iCLIP of mouse NXF1 and SRSF1/2/3/4/5/6/7	Müller-McNicoll et al.[63]	GEO: GSE69689
iCLIP of human NXF1, ALYREF, and CHTOP	Viphakone et al.[64]	GEO: GSE113953
mRNA half-lives in HeLa cells	Tani et al.[18]	Tables S1 and S2 in their paper
Experimental models: Cell lines		
Human/HeLa	gift from C.-H. Chung at 580 Seoul National University	ATCC
Oligonucleotides		
Oligo and spike-in sequences	This paper	See Table S6
Software and algorithms		
Scripts for data analysis	This paper	https://doi.org/10.5281/zenodo.10688611
ProteoWizard	Adusumilli and Mallick[87]	version 3.0.1908
MzXML2Search	Part of the Trans-Proteomic Pipeline (TPP)	version 5.2.0

(*Continued on next page*)

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MS-GF+	Kim and Pevzner[88]	version 2020.07.02
ProteomeDiscoverer	Thermo Fisher Scientific	version 2.4
python	https://www.python.org/	version 3.8.5
The scipy python package	https://scipy.org/	version 1.4.1
The numpy python package	https://numpy.org/	version 1.23.5
The pandas python package	https://pandas.pydata.org/	version 1.1.3
The scikit-learn python package	https://scikit-learn.org	version 0.23.2
R version 3.6.3	https://www.r-project.org/	version 3.6.3
The DEqMS R package	Zhu et al.[89]	version 1.4.0
The EmpiricalBrownsMethod R package	Poole et al.[90]	version 1.14.0
R version 4.1.0	https://www.r-project.org/	version 4.1.0
The org.Hs.eg.db R package	https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html	version 3.14.0
The topGO R package	Alexa et al.[25]	version 2.46.0

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, V. Narry Kim (narrykim@snu.ac.kr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Published datasets used in the study are provided in the key resources table under Deposited Data. All LC-MS3 data used in this study is available at the PRIDE database under accession number: PXD039054. RNA-seq data are available at Gene Expression Omnibus (GEO) under accession number GSE256124. The original images of the study are publicly available at https://doi.org/10.5281/zenodo.10688611.
- The data analysis codes are available at https://doi.org/10.5281/zenodo.10688611.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All cell lines used in this study tested mycoplasma-negative. HeLa (gift from C.-H. Chung at 580 Seoul National University) was authenticated by ATCC (STR profiling) and were grown in DMEM (Welgene, LM 001-05) containing 10% fetal bovine serum (FBS [Welgene, S001-01]).

METHOD DETAILS

Capture of the pulse-chased RNA interactome

HeLa cells were maintained in DMEM (Welgene) supplemented with 9% FBS (Welgene, cat#S001-01) and cultured at 37°C with 5% CO2. Four 150mm dishes were used for each time point sample. For pulse-labeling, cells were incubated with 0.5mM 4-thiouridine (Sigma-Aldrich, cat#T4509) for 10 min as suggested in previous papers.[91,92] After washing three times with cold PBS, cells were incubated with DMEM supplemented with 1mM uridine until UV irradiation. A "No-4sU" sample was used as a negative control. After incubation with the uridine supplemented DMEM for a certain designated incubation time (0-5 h), the pulse-chased cells were washed with cold PBS and irradiated with 365nm UV for 0.45J/cm2. After crosslinking, cells were immediately harvested and snap frozen in liquid nitrogen, and kept at −80 °C until further processing. Cell pellets were treated with TURBO DNase solution (160 Units per 150 mm dish, 1X TURBO DNase buffer in PBS) at 37°C for 30 min following the protocol in Lee et al.[51] The 2X lysis buffer (40mM Tris-HCl pH 7.5, 1M LiCl, 1% LiDS wt/vol, 1% NP40 wt/vol, 2mM EDTA, 10mM DTT, 8M Urea) were added to lyse the DNase treated cells. Lysates were homogenized by passing the lysate with a 21G needle. After measuring the amount of total protein with BCA assay, samples containing equal amounts of proteins (15mg) were used for RNA interactome capture (RIC). RIC

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

was performed as in previous papers[5] with the following modifications. Oligo(dT) bead (NEB, cat#S1419S), which was washed with lysis/binding buffer (20mM Tris-HCl pH 7.5, 500mM LiCl, 0.5% LiDS wt/vol, 0.5% NP40 wt/vol, 1mM EDTA, 5mM DTT, 4M Urea), was added to the lysate and incubated at room temperature for an hour. Beads were collected with the magnet, and the supernatant ("flowthrough") was transferred to a new tube and stored at 4°C for the additional capture (see below). The collected beads were washed once with lysis/binding buffer followed by two washes with wash buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.1% LiDS wt/vol, 0.5% NP40 wt/vol, 1 mM EDTA, and 5 mM DTT), wash buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.5% NP40 wt/vol, 1 mM EDTA, and 5 mM DTT), and wash buffer 3 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA and 5 mM DTT). All wash steps were performed at room temperature using pre-chilled wash buffers. For elution, beads were resuspended in 300 μl of elution buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA) and incubated for 3 min at 65°C with agitation before the supernatant containing eluted proteins was transferred to a fresh tube and stored. The elution step was repeated. The beads and the flowthrough from the first capture step were mixed to capture the residual RNPs remaining in the flowthrough. The incubation, wash, and elution steps were repeated, and the eluted proteins were combined with the eluted proteins from the first round.

Peptide sample preparation and TMT labeling for quantitative proteomics analysis

RNA-binding proteome samples collected via oligo(dT) bead pull down were first concentrated to 80 μl using speed-vac (Concentrator plus, Eppendorf) and then reduced with 20 mM dithiothreitol (DTT) (Sigma-Aldrich) at 37°C for 1h. The samples were placed onto 30kDa molecular weight cut off (MWCO) filter (Amicon 30kDa, Merck Millipore) along with the 200 μl of urea buffer (8 M urea in 25 mM HEPES buffer, pH 8.5), followed by centrifugation at 15,000×g for 15 min at 24°C. Each sample was then reconstituted with 200 μl of the urea buffer and centrifuged again (twice). The samples were then alkylated with 200 μl of 80mM iodoacetamide (Sigma-Aldrich) in the urea buffer and incubated at 37°C for 1 hour in the dark, followed by centrifugation. Each sample was then washed with 200 μl of the urea buffer twice and with 200 μl of 25 mM HEPES buffer (pH 8.5) twice. 100 ng of trypsin (~1:50 w/w, based on the estimation from TIC area) in 200 μl of HEPES buffer was added to each sample and incubated at 37°C for overnight. Samples were then centrifuged and the collected flow-throughs were concentrated to 40 μl using the speed-vac. Total of 11 samples were labeled with the TMT11plex reagents following the protocol provided by the manufacturer (Thermo Fisher). TMT labeled samples were combined and desalted using the C18 SPE cartridge (Supelco) and the elute from the cartridge was completely dried using the speed-vac and reconstituted with 50 μl of 10 mM ammonium bicarbonate (ABC) buffer.

A concatenated mid-pH (pH 8) RPLC off-line fractionation was carried out at micro-scale for multidimensional LC-MS3 analysis to improve the quantitative profiling depth. For micro-scale fractionation, a RPLC capillary column (320 μm i.d. x 55 cm) was in-house packed with Jupiter C18 beads (Phenomenex, 3 μm). The 50 μl of combined TMT11-labeled peptide sample was loaded onto the capillary column. A linear gradient of solvent A (10 mM ABC in water, pH 8) and solvent B (10 mM ABC in 90% acetonitrile) was applied on nanoAcquity (Waters) at a flow rate of 7 μL/min; 2% solvent B isocratic for initial 14 min, 2 to 10% solvent B for following 2 min, 10 to 40% solvent B for next 56 min. The eluent was automatically concatenated into 6 fractions using TriVersa NanoMate (Advion) and reconstituted with 25 mM ABC buffer for further LC-MS3 analysis.

Liquid chromatography and tandem mass spectrometry analysis

The TMT11-labeled 6 fractions were analyzed using an Orbitrap Eclipse via MS3 mode (Thermo Fisher Scientific) coupled with nanoAcquity UPLC system (Waters), which was equipped with an in-house packed trap (150 μm i.d. x 3 cm) and analytical column (75 μm i.d. x 100 cm) using 3 μm of Jupiter C18 particle (Phenomenex). During the analysis, the analytical capillary column was heated at 45°C with the column heater (Analytical Sales and Services). A linear gradient of solvent A (water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid) was applied at a flow rate of 300 nl/min as follows: 5 to 8% solvent B for initial 10 min, 8 to 35% solvent B for next 195 min. Total run time for the SPS-MS3 analysis were 220 min with the following set up for MS acquisition; Full MS scans (m/z 400–1600) were acquired at a resolution of 120k (at m/z 200) with 4E5 of AGC target value and 50 ms of ITmax. Selected precursor ions were first isolated at 0.7 Th of isolation window and subjected to HCD fragmentation for MS2 scans in orbitrap at a resolution of 15k (ITmax 60 ms, AGC 5E4 and NCE 30%). The 10 most intense MS2 fragment ions were synchronously isolated in ion trap for final HCD MS3 scans at a resolution of 50k and 0.4 Th of isolation width (AGC 13E5, ITmax 150 ms, and NCE 65%). Overall 3 s of cycle time was applied. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al.[93]) partner repository with the dataset identifier PXD039054.

Total RNA analysis

For comparison between conventional RIC and our modified RIC method described above, one 150mm dish of HeLa cells were harvested and lysed with RIC lysis/binding buffer (20mM Tris-HCl pH 7.5, 500mM LiCl, 0.5% LiDS wt/vol, 0.5% NP40 wt/vol, 1mM EDTA, 5mM DTT). After the RIC procedure described above, RNA was isolated from the elution. For total RNA isolation from inputs, 1 ml of TRIzol (Invitrogen) was added to 2.5% of input lysates (which is about 50 ul). The rRNA ratio was estimated by using Tapestation RNA screen tape.

Measuring 4sU labeling efficiency

HeLa cells were pulse-labeled with 0.5 mM 4sU for 10 min and used directly for RNA purification to obtain total RNA population or used for oligo(dT) enrichment. During RNA purification with TRIzol, DTT was added to aqueous phase in the final 0.1 mM and total

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

RNA was eluted in 1 mM DTT solution to prevent oxidation of the 4sU labeled RNA. For oligo(dT) enriched RNA samples, 5 mM DTT was added to the elution buffer after RIC. The eluted RNA was digested and dephosphorylated to single nucleosides with bacterial alkaline phosphatase (Invitrogen, final 1.6U/μl) and Phosphodiesterase I (Worthington, final 0.2 U/μl) as previously described in Herzog et al.[94]

Nucleoside samples prepared from each experimental condition was reconstituted in 50 μl of solvent A (200 mM triethylammonium acetate) and loaded onto a BEH C18 column (2.1 mm i.d. x 300 mm, 1.7 μm particle) (Waters) coupled with the 1290 Infinity UHPLC system (Agilent) and the column heater was set at 25°C. A linear gradient of solvent A and solvent B (90 % methanol) was applied at a flow rate of 100 μl/min as follows: 100% isocratic solvent A for initial 5 min, 0 to 20% solvent B for 20 min, 20 to 30% solvent B for 20 min, 30 to 90% solvent B for 10min, and 90 to 0% solvent B for 2 min, followed by isocratic 100% solvent A for 33 min. UV absorbance at 260 nm and 330 nm were monitored. U and 4sU standards with concentration of 8 μM and 800 μM, respectively, were either analyzed individually or as a mixture of 1:1 ratio (v/v) to determine the respective retention times of U and 4sU. Relative amount of U and 4sU in nucleoside sample prepared for each experimental condition was estimated based on the UV signal, area under the curve, obtained by 260 nm and 330 nm UV detection at the respective retention times of U and 4sU.

Conventional protein analyses

RIC samples were resolved on SDS-PAGE and analyzed by silver staining using EzWay Protein-Silver Staining Kit (KOMABIOTECH, cat#K14040D). For western blot analysis, the eluates of RIC were first concentrated with Amicon 30K Ultra-0.5 (Millipore) and treated with RNase A (Thermo Scientific, cat#EN0531) and Benzonase (Sigma-Aldrich, cat#E1014). The concentrated samples were then loaded on 10% Novex WedgeWell Tris-Glycine Mini Gel (Invitrogen). After transferring to a methanol-activated PVDF membrane (Millipore), the membrane was blocked in PBS-T containing 5% milk, probed with primary antibodies, and washed three times. Anti-mouse or anti-rabbit HRP-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories) were incubated and washed three times again. Chemiluminescence was performed with West Pico Luminol reagents (Thermo), and the signals were detected by ChemiDoc XRS+ System (BioRad). In detail, a sample was divided into three equal parts, and three gels were run for western blotting analysis to examine bands of similar sizes. LARP1 (Bethyl Cat# A302-087A, RRID:AB_1604274), ALYREF (Bethyl Cat# A302-892A, RRID:AB_10663772), TDP-43 (Proteintech Cat# 10782-2-AP, RRID:AB_615042) and G3BP1 (BD Biosciences Cat# 611126, RRID:AB_398437) were first detected, then the membrane was stripped with WSE-7240 EzReprobe (2332530) and restained with antibodies against hnRNPA1 (gift fromGideon Dreyfuss), SRSF7 (MBL International Cat# RN079PW, RRID:AB_11161213) and FMR1 (MBL International Cat# RN016P, RRID:AB_1953044). Additionally, PABPN (Abcam Cat# ab75855, RRID:AB_1310538), PABPC4 (Thermo Fisher Scientific Cat# A301-466A, RRID:AB_999661), POLR2 (Santa Cruz Biotechnology Cat# sc-56767, RRID:AB_785522), eIF4G2 (MBL International Cat# RN003P, RRID:AB_1570636) were first detected, then the membrane was stripped and restained with IGF2BP3 (Santa Cruz Biotechnology Cat# sc-365640, RRID:AB_10847223). NXF1 (Abcam Cat# ab129160, RRID:AB_11142853), EIF4A3 (Proteintech Cat# 17504-1-AP, RRID:AB_2097393) were detected on the last gel.

Immunofluorescence and confocal microscopy

HeLa cells were cultured on 8-chamber slide glass (Thermo, 154461PK). The cells were labeled for 10 min with 4sU, followed by a 5 hrs chase with uridine-supplemented media, as previously described. To induce cellular stress, cells were treated with 0.5mM sodium arsenite for 1 hour. Subsequently, the cells were fixed using 4% paraformaldehyde for 10 min and permeabilized with 0.1% Triton X-100. After blocking with a 5% BSA buffer, the cells were incubated with anti-G3BP1 antibody (BD Biosciences Cat# 611126, RRID:AB_398437) for 2 hours at room temperature. Following washing steps, the cells were incubated with an anti-mouse Alexa Fluor 488-conjugated secondary antibody and DAPI (D9542) for 1 hour. Images were captured using a Nikon ECLIPSE Ti2 microscope.

LC-MS3 data processing for peptide identification

For all proteome data analysis, canonical protein sequences (SwissProt) of UniProt human reference proteome UP000005640 (last modified on Dec 3th, 2018, 20303 proteins were included) was used. Pig trypsin (UniProt accession P00761) and the cRAP (common Repository of Adventitious Proteins) protein sequences version 2012.01.01 (http://www.thegpm.org/crap/) were appended to the search space, to mark the peptides from the common contaminant proteins. To estimate peptide level FDR, decoy sequences were generated by reversing target protein sequences. The RAW format LC-MS3 data file was first converted into mzXML format using msconvert (ProteoWizard version 3.0.1908, Adusumilli et al.[87]) with the following parameters: –ignoreUnknownInstrumentError –mzXML –filter "peakPicking true [1,2]" –filter "msLevel 1-2". Then, peak count filtering ($\geq$ 20 peaks), charge suggestion (within range +1 - +7) and mzXML to MGF conversion was done by MzXML2Search (v5.2.0) with the following parameters: -mgf -B0 -T20000 -c1-7 -P20. Resulting MGF files were subjected to the MS-GF+ (v2020.07.02, Kim et al.[88]) for peptide identification at the search tolerance 10ppm. For MS-GF+ search, we set carbamidomethylation of cysteine and TMT labeling of peptide N terminus/Lysine as static modification, and oxidation of methionine and acetylation of protein N terminus as variable modification. Also, -m 3 option was set for the HCD fragmentation.

Peptide to protein grouping and assigning MS3 quantities to protein groups

Protein groups were generated from the identified MS2 scans by inhouse script, which implements the greedy algorithm described in Zhang et al.[95] Protein group level FDR was calculated using decoy proteins, with formula N_{decoy} / N_{target}. To obtain

the TMT label quantities from MS3 scans, RAW files were processed with ProteomeDiscoverer (version 2.4) with isotope impurity correction for TMT labels (Thermo Fisher Scientific TMT reagent, lot number UL292365 and UK288917), by setting "Apply Quan Value Corrections"=True in "reporter ions quantifier" node. Low quality MS3 scans were discarded with SNR $\geq$ 5 and isolation interference < 70% cutoffs. MS3 scans analyzed by ProteomeDiscoverer were mapped to the MS-GF+ identified PSMs, by inhouse script comparing the scan number (MS2 scan number < MS3 scan number) and the precursor ion m/z (within 1E-9 Th). Quantity of a protein group was calculated by adding up quantities of all matched PSMs, including both unique and razor PSMs.

Protein quantity processing and contaminant filtering

We carried out the following procedures to generate each protein's quantity dynamics from triplicate experiments. First, protein quantities were normalized against the quantity of pig trypsin (UniProt accession P00761), which was equally added to each sample during sample preparation. Subsequently, to remove the background from non-specific binders, the protein intensity from each channel was subtracted by that from the No4sU control channel. Pseudo intensity was set as the bottom 5% quantity of each replicate and added to all peptides' intensities to avoid division by zero. To ensure each replicate contributes the same to the final protein quantity, protein intensities of each replicate were normalized so that the intensity sum of a protein in each replicate is the same with the median of summed intensity of single replicate in all proteins. Normalized and $\log_2$-transformed replicate quantities were merged and smoothed by univariate quadratic spline (using scipy version 1.4.1, https://scipy.org/), using time points as independent variables and quantities as dependent variables. The univariate spline was unsuccessful on 18 proteins with very high variance among replicates, so those proteins were discarded. The peak binding time of RBP was found from the fitted spline line, interpolated at 0.1 min interval. Reproducibility of protein quantification was defined as the ratio between per replicate variation and per time-point variation. In practice, it was calculated as root mean squared error (RMSE) between measured quantities and merged protein quantity spline (per replicate variation) divided by standard deviation (std) of merged protein quantities (per time-point variation). Proteins with RMSE / std ratio above 1 were discarded.

Before further analyzing data, proteins marked as common lab contaminants by cRAP database were discarded. Also, human keratin and histone proteins were removed before the analysis, because those proteins are less likely to act as real RBP, and showed high variance between replicates.

Statistical enrichment analysis for mRNP capture

To determine the enrichment of RBPs in captured mRNPs, we compared labeled samples from 10 time points with the non-labeled ("No-4sU") sample. This comparison was performed iteratively for a total of 10 times and the results were subsequently combined, with the following procedure. Initially, protein quantities were normalized and subjected to a $\log_2$ transformation. The DEqMS package[89] was then utilized to calculate the differential expression p-value for each labeled sample versus the No-4sU sample pair. Next, the 10 individual p-values were merged into a single p-value using the empirical Brown's method.[90] Considering the large number of proteins ($\sim$1,000) tested, multiple test correction was applied to the merged p-values of each protein using the Benjamini-Hochberg method. Proteins that had an adjusted P value < 0.01 and were more abundant in the labeled sample than in the No-4sU sample were defined as enriched RBP.

Previously identified mRNP list from RBP2GO

We adopted previously identified mRNP list from the meta analysis database, RBP2GO.[9] All human RBP datasets were taken, except for *in vitro* experiments, *in silico* prediction, review or meta analysis, and non-poly(A) enrichment datasets. The RBP2GO identifiers of used datasets are following: Baltz_HEK293_2012, Castello_HeLa-S3_2012, Beckmann_HuH-7_2015, Castello_HeLa-S3_2016, Conrad_K562_2016, Milek_MCF7_2017, Perez-Perri_Jurkat_RIC_2018, Perez-Perri_Jurkat_eRIC_2018, Garcia-Moreno_HEK293_2019, Backlund_HuH-7_Cytoplasmic_2020, Backlund_HuH-7_Nuclear_2020, Kramer_HeLa_2014, Panhale_HEK293_2019, Mullari_HEK293_2017.

Protein domain enrichment analysis

Taxon 9606 (human) protein domain annotations in Pfam database (version 32.0) was used for protein domain enrichment analysis. One-sided Fisher's exact test was applied to estimate the statistical enrichment of a particular domain among the quantified RBPs. Benjamini-Hochberg method was applied to the p-values for the multiple test correction.

Clustering analysis on temporal dynamics of RBPs

For the feature standardization, log-transformed quantities were z-score transformed for protein-wise direction, by the mean and standard deviation of all time points for each protein. Also, the peak binding time of each protein was z-score transformed at feature-wise direction, by the mean and standard deviation of peak binding time of all RBPs. K-means clustering using above normalized features was done by scipy version 1.4.1 (https://scipy.org/). The mean distortion of clustering result was calculated as the mean Euclidean distance between the feature values and the cluster centroids. The number of clusters were manually set to 7 after testing several cluster numbers ($5 \leq K \leq 10$), based on the mean distortion of clustering results.

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Filtering out non-protein-coding RNA binders and mitochondrial mRNA binders

Several RBPs in our data are primarily known as non-coding RNA or mitochondria-coded mRNA binders, even after we filtered RBPs with previously reported mRNP list. Thus, we defined annotation and knowledge based blacklist for the further analyses. First of all, mitochondrial mRNA binders were defined as genes annotated with GO "mitochondrial matrix" but not with GO "cytosol". Then, we added ribosomal proteins and mitochondrial ribosome proteins to the blacklist. Next, GO "pre-snoRNP complex", "sno(s)RNA-containing ribonucleoprotein complex", and "preribosome" annotated genes were added to the blacklist. Lastly, three proteins (DCAF13, FTSJ3, NGDN) were removed from the blacklist and re-included to the mRNP list, as there are literatures supporting their binding on nuclear-coded mRNAs. "Ribosome or mitochondria related blacklist" column in Table S2 shows the list of proteins filtered out by this process.

Pulse-chased RNA sequencing

One 150 mm culture dish of HeLa cell for each time point was used for the experiment. Followed by 10 min of 4-thiouridine labeling, cells were harvested after incubation for 0 min, 30 min, and 5 hrs of uridine supplemented media. As a negative control, a no-4-thiouridine-labeled sample ("No-4sU") was also prepared. Cells were harvested without UV crosslinking and then treated with DNaseI. For spike-in RNAs, *in vitro* transcribed non-human RNAs were prepared with or without 4-thiouridine and then polyadenylated with ePAP (NEB). After adding spike-in RNAs to the same amount of lysate, the RIC protocol described above was used to enrich poly(A) RNA in the lysate except elution buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, and 1mM DTT). The eluted samples were precipitated with ethanol. Enrichment of 4-thiouridine labeled RNA was performed based on the previously described protocol[96] with the following alterations. 50ug of poly(A) enriched RNAs were used for MTS-pull-down. MTS-biotin was diluted in N,N-Dimethylformamide (Sigma, cat#227056). RNAs were incubated with diluted MTS-biotin at room temperature for 2 hrs and purified with RNAclean XP (Beckman, cat#A63987). Dynabeads MyOne Streptavidin C1 (Invitrogen, cat#65002) were washed with nuclease-free water and high salt buffer (10 mM Tris, 1 mM EDTA, 100 mM NaCl, 0.05 % Tween-20 wt/vol) and then blocked with blocking buffer (10 mM Tris, 1 mM EDTA, 100 mM NaCl, 0.05 % Tween-20 wt/vol, 5 ug/ul glycogen) for an hour at room temperature. Biotinylated RNAs were heated at 65°C for 10 min and placed on ice. 10X High salt buffer was added to RNA solution to make it 1X concentration, and pre-washed beads were added to RNAs. Biotinylated RNA and streptavidin beads were incubated in dark for 15 min. The first supernatants were kept as flow-through. After washing beads with high salt buffer three times, 4-thiouridine labeled RNAs were eluted from beads with elution buffer (100 mM DTT, 5% 2-Mercaptoethanol, 20 mM HEPES pH 7.6, 1 mM EDTA, 100 mM NaCl, 0.05% Tween-20 wt/vol). RNA-seq libraries were constructed using MGIEasy RNA Directional Library Prep Kit V2.0 (MGI, cat#1000006385) and sequenced by paired-end run on MGI sequencer.

RNA-seq sequence processing and alignment

The initial parts of sequence analysis were done by using Cutadapt version 3.0.[97] For 5′ and 3′ end of each read, the low-quality bases below Phred quality of 30 were trimmed. After trimming, 3′ adaptor sequences of both the first and second read of each pair were removed (Table S6). Read pairs with any read shorter than 70 bases were removed after trimming and adopter clipping. To get the readcount for each gene and spike-in, a transcript reference was build based on spike-in sequences (Table S6), UCSC Genome Browser hg38 RefGene annotation (downloaded on February 20, 2020), and hg 38 genome, by RSEM (v1.3.1).[98] Pre-mRNA transcript models were also generated and added to the transcript reference by in-house software, since un-spliced mRNAs might be included in the early stage mRNPs. The read pairs were aligned to the above transcript reference by STAR version 2.7.6a.[99] The read count of each gene and spike-in was calculated by RSEM (v 1.3.1).[98] Spike-in normalization was done by the ratio of geometric mean of spike-in TPMs.

Normalizing RNA-seq read counts to estimate 4sU labeled mRNA counts in pulse-chased RNA sequencing results

In the oligo(dT) + MTS-biotin pull-down RNA-seq samples, we observed non-negligible amounts of 4sU-unlabeled RNA spike-ins. This indicates that MTS-biotin pull-down method does not fully remove the 4sU-unlabeled mRNAs. Thus, we calculated the portion of 4sU-unlabeled mRNAs and subtracted that portion, by following two steps: 1) normalizing RNA-seq read count values using 4sU-unlabeled spike-ins, and 2) subtracting the normalized read count of oligo(dT) capture-only sample from the normalized read count of oligo(dT) + MTS-biotin labeling and pull-down sample. We justify this procedure as follows.

Let the observed read count value of the transcript t in sample x enriched by enrichment method e (MTS or dT) as $O_{t\,x\,e}$. Let the true RNA amount of the same transcript in the same sample with 4sU labeling status l (4sU or No4sU) as $T_{t\,x\,l}$. Also, let the efficiency of the oligo(dT) capture-only and oligo(dT) + MTS-biotin labeling and pull-down in sample x for labeling status l as $E_{x\,l}$ and $F_{x\,l}$, respectively. Then, following relationships between observed read count values and the true RNA amount can be formulated.

Observed read count and true RNA amount of transcript t in sample x, oligo(dT) + MTS-biotin labeling and pull-down library:

$$O_{t\,x\,MTS} = E_{x\,MTS}T_{t\,x\,No4sU} + F_{x\,MTS}T_{t\,x\,4sU}$$

$$\Rightarrow T_{t\,x\,4sU} = \frac{1}{F_{x\,MTS}}\left(O_{t\,x\,MTS} - E_{x\,MTS}\,T_{t\,x\,No4sU}\right) - \qquad \text{(Equation 1)}$$

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Observed read count and true RNA amount of transcript t in sample x, oligo(dT) capture only library:

$$O_{t\,x\,dT} = E_{x\,dT}\left(T_{t\,x\,No4sU} + T_{t\,x\,4sU}\right)$$

$$\Rightarrow T_{t\,x\,No4sU} = \frac{1}{E_{x\,dT}} O_{t\,x\,dT} - T_{t\,x\,4sU} - \quad \text{(Equation 2)}$$

By substituting $T_{t\,x\,No4sU}$ equation in Equation 1 by Equation 2,

$$T_{t\,x\,4sU} = \frac{E_{x\,MTS}}{F_{x\,MTS}\left(\frac{1}{E_{x\,MTS}} O_{t\,x\,MTS} - \frac{1}{E_{x\,dT}} O_{t\,x\,dT} + T_{t\,x\,4sU}\right)}$$

$$\Rightarrow T_{t\,x\,4sU} = \frac{1}{\left(\frac{F_{x\,MTS}}{E_{x\,MTS}} - 1\right)\left(\frac{1}{E_{x\,MTS}} O_{t\,x\,MTS} - \frac{1}{E_{x\,dT}} O_{t\,x\,dT}\right)} \quad \text{(Equation 3)}$$

Here, the ratios of sequencing efficiency $E_{x\,l}$ and $F_{x\,l}$ can be obtained from the spike-in read counts. Let the observed read count value of the spikie-in with 4sU labeling status l in sample x enriched by enrichment method e (oligo(dT) only or oligo(dT) + MTS-biotin labeling and pull-down) as $P_{x\,e\,l}$. Let the true RNA amount of the same spike-in in the same library as $U_{x\,e\,l}$. Then, following relationships between spike-in read counts and sequencing efficiencies can be established. Of note, many species of spike-in were included in practice, so the geometric mean of all spike-in read counts was used as the observed read count $P_{x\,e\,l}$.

4sU labeled spike-ins in sample x, oligo(dT) only:

$$P_{x\,dT\,4sU} = E_{x\,dT} U_{4sU}$$

4sU labeled spike-ins in sample x, oligo(dT) + MTS-biotin labeling and pull-down):

$$P_{x\,MTS\,4sU} = F_{x\,MTS} U_{4sU}$$

Non-4sU labeled spike-ins in sample x, oligo(dT) only:

$$P_{x\,dT\,No4sU} = E_{xdT} U_{No4sU}$$

Non-4sU labeled spike-ins in sample x, oligo(dT) + MTS-biotin labeling and pull-down):

$$P_{x\,MTS\,No4sU} = E_{x\,MTS} U_{No4sU}$$

Using the above equations, the relationship between sequencing efficiency value $E_{x\,l}$ and $F_{x\,l}$ can be formulated as following:

$$\frac{1}{E_{x\,dT}} = \frac{U_{No4sU}}{P_{x\,dT\,No4sU}} - \quad \text{(Equation 4)}$$

$$\frac{1}{E_{x\,MTS}} = \frac{U_{No4sU}}{P_{x\,MTS\,No4sU}} - \quad \text{(Equation 5)}$$

$$F_{x\,MTS} = \frac{P_{x\,MTS\,4sU}}{U_{4sU}} - \quad \text{(Equation 6)}$$

$$\frac{U_{No4sU}}{U_{4sU}} = \frac{P_{x\,dT\,No4sU}}{E_{x\,dT}} \cdot \frac{E_{x\,dT}}{P_{x\,dT\,4sU}} = \frac{P_{x\,dT\,No4sU}}{P_{x\,dT\,4sU}} - \quad \text{(Equation 7)}$$

$$\frac{F_{x\,MTS}}{E_{x\,MTS}} = \frac{P_{x\,MTS\,4sU}}{U_{4sU}} \cdot \frac{U_{No4sU}}{P_{x\,MTS\,No4sU}} = \frac{P_{x\,MTS\,4sU}}{P_{x\,dT\,4sU}} \cdot \frac{P_{x\,dT\,No4sU}}{P_{x\,MTS\,No4sU}} - \quad \text{(Equation 8)}$$

By applying Equations 4, 5, 6, and 8 to Equation 3:

$$T_{t\,x\,4sU} = \frac{1}{\frac{P_{x\,MTS\,4sU}}{P_{x\,dT\,4sU}} \cdot \frac{P_{x\,dT\,No4sU}}{P_{x\,MTS\,No4sU}} - 1} \left(\frac{O_{t\,x\,MTS}}{P_{x\,MTS\,No4sU}} - \frac{O_{t\,x\,dT}}{P_{x\,dT\,No4sU}}\right) U_{No4sU}$$

Please cite this article in press as: Choi et al., Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.012

Since the input amount of the spike-in among all libraries are the same, U_{No4sU} could be considered as a constant. Also, $\frac{P_{x\ MTS\ 4sU}}{P_{x\ dT\ 4sU}} \cdot \frac{1}{\frac{P_{x\ dT\ No4sU}}{P_{x\ MTS\ No4sU}} - 1}$ part is only affected by the sample x and capture method, but not changed by the transcript t. Hence, for the proportion of 4sU labeled mRNAs of transcript t in sample x in MTS-biotin captured libraries, the below proportional expression is established.

$$T_{t\ x\ 4sU} \propto \left(\frac{O_{t\ x\ MTS}}{P_{x\ MTS\ No4sU}} - \frac{O_{t\ x\ dT}}{P_{x\ dT\ No4sU}} \right)$$

Thus, to obtain values proportional to 4sU mRNA quantities, we can simply normalize the mRNA read count using the 4sU-unlabeled spike-ins, then subtract normalized read count of oligo(dT) capture-only from the normalized read count of oligo(dT) + MTS-biotin labeling and pull-down.

Gene Ontology and subcellular localization analysis

As previously described, ribosomal proteins, snoRNA related factors, and RBPs exclusively localized to the mitochondrial matrix were removed in this analysis. All GO gene annotations used in this study were obtained from the org.Hs.eg.db R package (version 3.14.0). To gain functional insights to each cluster, GO term enrichment analysis was done by Fisher's exact test, by setting all 1035 proteins identified by MS as the statistical background. P-values were adjusted for GO hierarchy and local dependencies by *weight01* algorithm in topGO R package.[25] To test GO terms enriched at a specific time point, we utilized pairwise distances between RBPs used in prior clustering analysis. Per each GO term, pairwise distances within RBPs annotated with GO term and not annotated with GO term were compared by Mann-Whitney U test. To remove redundantly enriched parent-child GO terms, the *elim* algorithm (described in Alexa et al.[25]) was applied to adjusted p-values. To classify nuclear/cytoplasmic RBPs, we utilized image-based data from Human Protein Atlas (downloaded March 8, 2019) and combined localizations with reliability of Enhanced, Supported, or Approved. For deeper analysis of protein subcellular localizations, we downloaded Human cell map database v1[29] which provides two sets of predictions, from the SAFE and the NMF algorithm. We primarily used SAFE results and supplemented them using NMF, as described in Lee et al.[51] Detailed procedures for merging SAFE and NMF results can be found in the deposited source code (https://doi.org/10.5281/zenodo.10688611).

Binding site mapping of RBPs

As previously described, ribosomal proteins, snoRNA related factors, and RBPs exclusively localized to the mitochondrial matrix were removed in this analysis. The binding locations of RNA-binding proteins (RBPs) were determined based on eCLIP peak locations provided by the ENCODE project.[30–32] To generate non-redundant transcript annotations, we initially calculated read counts for duplicated RNA-seq data from HepG2 and K562 cell lines using RSEM, as previously described (ENCODE data accession: ENCFF002DKZ, ENCFF002DLC, ENCFF002DLE, ENCFF002DLG for HepG2; ENCFF001RDE, ENCFF001RCW, ENCFF001RDD, ENCFF001RCV for K562). Subsequently, we selected the most abundant transcript for each protein-coding gene, based on TPM values calculated by RSEM. The transcript regions were then divided into subregions, including introns, 5′ UTRs, CDS, and 3′ UTRs. Finally, for all overlapping subregions, we selected the subregion from the most abundant transcript. In cases where TPM values were tied, which was very rare, we prioritized genes with the smallest RefSeq accession number. The overlap between eCLIP peaks and transcript annotations was determined using the intersect tool in Bedtools, with a minimum requirement of $\geq$ 50% overlap relative to the span of the eCLIP peak.[100]

Protein-protein interaction network and protein complex analysis

As previously described, ribosomal proteins, snoRNA related factors, and RBPs exclusively localized to the mitochondrial matrix were removed in this analysis. We downloaded PPI data from the BioGRID database (release 4.2.191, Stark et al.[35]). Also, we considered only human PPIs (1) classified as physical interaction and (2) found from at least two different types of experiments, and (3) supported by at least two publications. The networkx python package was used for graph visualization. For the protein complex annotation, we utilized the CORUM database (release 3.0_03.09.2018_coreComplexes).

To quantitatively measure the differences among the mRNA binding dynamics of protein within PPI pairs or protein complexes, we calculated Euclidean distances between mRNA binding dynamics vectors, defined as [*log_2 protein intensity at 0 min, log_2 protein intensity at 15 min, …, log_2 protein intensity at 300 min, peak binding time*]. Before calculating Euclidean distances, protein intensities and peak binding times were normalized as described in "clustering analysis on temporal dynamics of RBPs" section.

mRNA binding dynamics prediction from GO annotations

As previously described, ribosomal proteins, snoRNA related factors, and RBPs exclusively localized to the mitochondrial matrix were removed in this analysis. To generate a Boolean table of RBPs versus GO, we utilized the org.Hs.eg.db R package (v3.14.0). Each cell in this table was marked as "true" if an RBP was annotated with a GO term, and "false" otherwise. To remove less informative annotations, we excluded GO terms with fewer than 7 annotated or unannotated RBPs. Subsequently, we performed multiple correspondence analysis (MCA) to convert the Boolean table into numeric values and reduce its dimensionality. From the MCA

transformed features, we selected the top 30% (323 features), which retained approximately 98% of the original Boolean table's information. As the presence of imbalanced numbers of early and late binder RBPs could introduce prediction accuracy bias, we mitigated this bias by oversampling RBPs. This involved duplicating certain RBPs in clusters with fewer members, ensuring that all clusters contained an equal number of RBPs (200). For each time point, we trained a ridge regression model to predict the z-scores of individual RBPs using the MCA transformed features. To determine the optimal model parameters, we performed 5-fold cross-validation. The scikit-learn python package (v0.23.2, https://scikit-learn.org) was utilized for fitting the regression models and conducting cross-validation.

QUANTIFICATION AND STATISTICAL ANALYSIS

Proteome and transcriptome quantification procedures are described in the method details section. Statistical analyses were conducted as outlined in figure legends and method details section, with additional details available in the deposited source code (https://doi.org/10.5281/zenodo.10688611).