

# An RNA tagging approach for system-wide RNA-binding proteome profiling and dynamics investigation upon transcription inhibition

Zheng Zhang<sup>1,†</sup>, Tong Liu<sup>1,†</sup>, Hangyan Dong<sup>1</sup>, Jian Li<sup>1</sup>, Haofan Sun<sup>1</sup>, Xiaohong Qian<sup>1</sup> and Weijie Qin<sup>1,2,\*</sup>

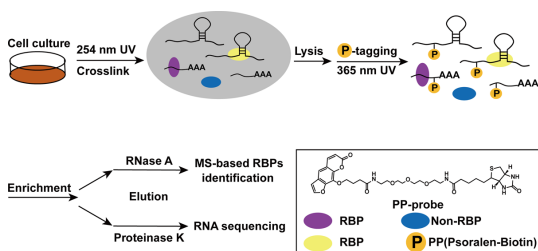
<sup>1</sup>National Center for Protein Sciences Beijing, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Lifeomics, Beijing 102206, China and <sup>2</sup>College of Basic Medicine, Anhui Medical University, Hefei 230032, China

Received June 05, 2020; Revised February 17, 2021; Editorial Decision February 24, 2021; Accepted March 08, 2021

## ABSTRACT

RNA-protein interactions play key roles in epigenetic, transcriptional and posttranscriptional regulation. To reveal the regulatory mechanisms of these interactions, global investigation of RNA-binding proteins (RBPs) and monitor their changes under various physiological conditions are needed. Herein, we developed a psoralen probe (PP)-based method for RNA tagging and ribonucleic-protein complex (RNP) enrichment. Isolation of both coding and noncoding RNAs and mapping of 2986 RBPs including 782 unknown candidate RBPs from HeLa cells was achieved by PP enrichment, RNA-sequencing and mass spectrometry analysis. The dynamics study of RNPs by PP enrichment after the inhibition of RNA synthesis provides the first large-scale distribution profile of RBPs bound to RNAs with different decay rates. Furthermore, the remarkably greater decreases in the abundance of the RBPs obtained by PP-enrichment than by global proteome profiling suggest that PP enrichment after transcription inhibition offers a valuable way for large-scale evaluation of the candidate RBPs.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Eukaryotic genomes encode a large number of RNA-binding proteins (RBPs) that interact with RNAs to form dynamic complexes to execute biological functions. These RBPs extensively control every stage of the life cycle of RNA (1,2), including the synthesis and maturation of mRNA, the activation of noncoding RNAs (ncRNAs) in chromatin remodelling and epigenetic regulation and the final decay of RNAs. Thus, RNA-protein interactions are increasingly considered a diverse and critical layer of transcriptome and proteome regulation (3,4). Considering their functional importance, the precise composition and dynamics of RBPs in ribonucleic-protein complex (RNP) under cell stress are receiving increasing attention. The disruption of RBPs may lead to cellular dysfunction and numerous diseases, including metabolic disorders, muscular atrophies, neurological disorders, autoimmune disease and cancer (5,6). Therefore, the systematic identification of RBPs and the quantification of their dynamic changes in RNPs is of great significance for understanding their diverse roles in regulating fundamental cellular functions and disease states (7). The comprehensive mapping of RBPs and the elucidation of their dynamic assembly in RNPs demand efficient tools for the specific and unbiased enrichment of RBPs.

Extensive efforts have been devoted to obtain a global view of RBPs and RNPs in recent years (8). Benefitting from UV irradiation-induced cross-linking that ‘freezes’ natural RNPs in intact cells, cross-linking and immunoprecipitation (CLIP)-based strategies have enabled the large-scale analysis of RBPs in different kinds of cells (9,10). Hundreds to thousands of RBPs have been identified by mass spectrometry analysis using enrichment strategies that exploit the polyA tails of mRNA (11,12), alkynyl uridine analog labelling (13,14), solubility of RNPs in organic solvents (15–17) or silica affinity (18,19). These methods largely promoted high-through profiling of RBPs and discovered po-

\*To whom correspondence should be addressed. Tel: +86 10 61777111; Fax: +86 10 61777050; Email: aunp\_dna@126.com

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

tential new functions of RBPs in regulating cellular activity. Despite the rapidly expanding RNA-binding proteome, there are still some limitations in the current methods. For alkynyl uridine analog-based RNPs enrichment, metabolic labelling may interfere the physiological process of the cell. Although do not involve using non-canonical uridine, the solubility- or silica affinity-based methods may face the glycoprotein contamination issue, due to the similar solubility or silica affinity between RNPs and glycoproteins (16,20–22). Therefore, we developed a photoreactive psoralen probe (PP) for the global enrichment of in vivo UV cross-linked RNPs (Scheme 1) and RBPs identification by mass spectrometry (MS). In this strategy, the specific isolation of RNPs is achieved by psoralen tagging of RNA upon 365 nm UV irradiation. Psoralens are a class of planar, three-ring heterocyclic compounds that can form covalent bonds through cycloaddition reactions with uridine in RNA upon irradiation with 320–410nm UV (23–25). Although DNA can also react with psoralen via thymine and be captured by streptavidin beads, UV crosslinking of protein to DNA is much less efficient than that between proteins and RNA (15,26) and requires energy doses 10–100 times higher than that used for protein-RNA crosslinking (27). Furthermore, RNase A treatment is adopted for elution of the enriched RBPs by RNA degradation, further eliminating contamination by DNA-binding proteins. By the cycloaddition reactions between psoralen and uracil, the comprehensive isolation of both coding and noncoding RNPs achieved without the need of metabolic labelling of RNAs. In total, 2986 RBPs, including 782 putative novel RBPs were identified by MS analysis after PP enrichment from HeLa cells. These proteins cover ~70% of the RBPs obtained from the same cell line by previously reported polyA-dependent and polyA-independent strategies in five different works (11,13–15,28). The 782 candidate RBPs include 178 metabolism related proteins distributed among a variety of pathways, indicating a possibly broader interplay between RNA and metabolism than previously anticipated, which deserves further exploration. Furthermore, the large-scale dynamic investigation of the PP-enriched RBPs in the RNPs during RNA transcription inhibition reveals distinct decreasing patterns compared with the corresponding RBPs in global proteome profiling, therefore, providing further evidence to support the candidate RBPs. This finding suggests the potential of this strategy as a high-throughput way for evaluation of the candidate RBPs. Considering the different enrichment mechanism of PP, phase-separation and silica affinity methods (covalent-binding/dissolution property/affinity adsorption), combined application of these complementary methods is advantage for achieving more comprehensive enrichment and deeper coverage of RNA-binding proteome.

## MATERIAL AND METHODS

### Synthesis of the psoralen probe (PP)

The synthesis steps and characterization of PP are shown in the Supplementary Figure S1A–C, Supporting Information.

### Cell and culture conditions

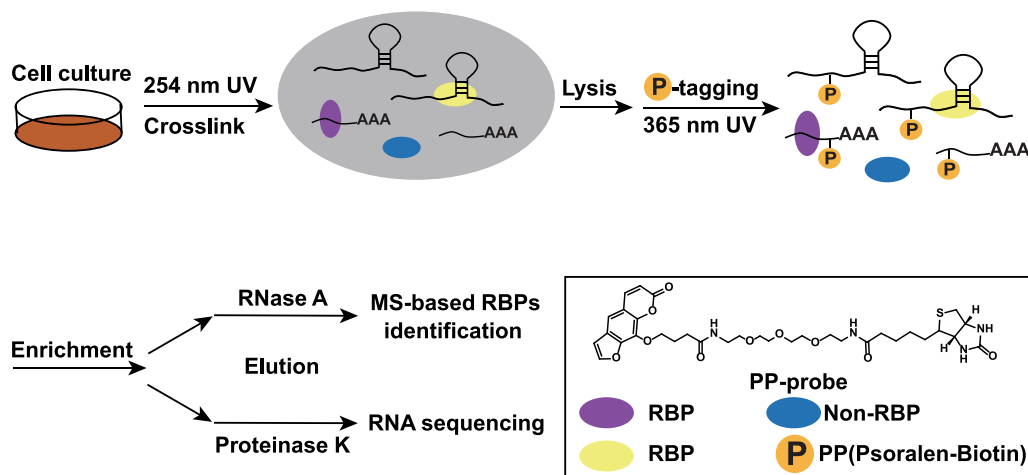
HeLa cells were grown in DMEM (Gibco) supplemented with 10% FBS (Gibco), 100 U/ml penicillin, and 100 µg/ml streptomycin (Gibco) at 37°C in a 5% CO<sub>2</sub> atmosphere. For the actinomycin D (ActD) treatment experiment, HeLa cells were treated with 5 µM ActD (J&K Scientific) at 37°C in DMEM medium (10% FBS, 100 U/ml penicillin, and 100 g/ml streptomycin) for 0, 2, 5 and 9 h before collecting the cells.

### Isolation of RNPs by in vivo cross-linking and RNA tagging by PP

All the buffers were prepared using RNase-free H<sub>2</sub>O. The HeLa cells were washed with 5mL cold PBS for three times, followed by irradiation with 254-nm UV light at 0.25 J/cm<sup>2</sup> on ice using a UV cross-linker (CL-1000; UVP). After adding 2 ml cold PBS, the cells were harvested using a cell lifter (Corning) and collected in a 1.5-ml RNase-free tube. After centrifugation and discarding the supernatants, the cells were re-suspended in 250 µl lysis buffer I [PBS, 0.5% SDS, EDTA-free protease inhibitor mixture (Thermo), ribonucleoside vanadyl complex (New England Biolabs)] and homogenized by passing through a narrow needle, followed by incubation at 4°C with gentle rotation for 20 min. All the buffers were prepared using RNase-free H<sub>2</sub>O. Next, the homogenate was adjusted with 1 ml lysis buffer II containing PBS, 0.1% Triton X-100, EDTA-free protease inhibitor mixture (Roche) and ribonucleoside vanadyl complex (New England Biolabs) and homogenized by passing through a narrow needle, followed by incubation at 4°C with gentle rotation for 20 min. The cell lysates was centrifuged at 16 000 g for 10 min and the supernatant was transferred to another 1.5 ml RNase-free tube. The supernatant was treated with 5 µM PP at 4°C for 20 min, followed by irradiation on ice with 365-nm UV light at 2 J/cm<sup>2</sup> using a UV cross-linker (CL-1000; UVP) for 4 min. The supernatant was concentrated to a final volume of 100 µl by an Amicon Ultra-2 centrifugal filter unit (molecular weight cutoff of 10 kDa, Millipore) and washed three times with 500 µl cold PBS. The PP treated lysate was adjusted to a volume of 500 µl with 2 M urea, and mixed with 100 µl precleared streptavidin magnetic beads (Thermo), followed by incubation with gentle rotation at 4°C for 1 h. After discarding the supernatant, the beads were washed twice with 200 µl PBS containing 0.2% SDS for 1 min, twice with 200 µl PBS containing 8 M urea for 1 min, twice with 200 µl PBS for 1 min and twice with 50 mM NH<sub>4</sub>HCO<sub>3</sub> buffer for 1 min. For RBPs elution, the beads were incubated with 20 µl 0.01 µg/µl RNase A at 37°C for 1 h. The eluted RBPs were analysed by SDS-PAGE with silver staining or MS-based proteomic analysis. For control samples in MS-based proteomic analysis, the cells and lysates were subjected to the same treatment, except that the beads were incubated with 0.1 µg/µl RNase A at 37°C for 1 h before the washing step and discarding the washing solution.

### Analysis of the PP captured RNA

*Preparation of the RNA sample.* The RNPs from HeLa cells were captured by streptavidin magnetic beads using the



**Scheme 1.** Schematic overview of the PP-based RNP enrichment for large-scale RNA sequencing and RBP identification by MS.

method as described above. Subsequently, the beads were resuspended in 400  $\mu$ l elution buffer (12.5 mM biotin, 75 mM NaCl, 7.5 mM Tris-HCl, pH 7.5, 1.5 mM EDTA, 0.15% SDS, 0.075% sarkosyl and 0.02% Na-deoxycholate dissolved in RNase-free H<sub>2</sub>O) and were incubated at RT for 20 min on a shaker, followed by heating at 65°C for 10 min. The solution was collected and the beads were eluted again to give 800  $\mu$ l solution in total. 800  $\mu$ l Proteinase buffer (100 mM Tris-HCl, pH 7.5, 12.5 mM EDTA, 150 mM NaCl and 2% SDS dissolved in RNase-free H<sub>2</sub>O) and 2 mg/ml Proteinase K (Ambion) was added and incubated at 55°C for 1 h. The eluted RNAs were further purified by TRIzol following the manufacturer's instructions. Next, the RNA samples were prepared with the TruSeq RNA Library Prep Kit v2 (Illumina, not stranded). The samples in total were barcoded to be sequenced on HiSeq2500.

**Analysis of RNA-Seq Data.** For estimation of the rRNA content of libraries, reads were aligned to a collection of human ribosomal sequences which retrieved from NCBI nucleotide database. All reads were aligned to those sequences using bowtie2 and reads that failed to align were saved as non-rRNA reads file. Reads in this file were aligned to the complete hg38 assembly by using STAR. Percentage of the rRNA content was estimated by comparing the number of reads aligning to the rRNA sequences and the residual reads aligning to the complete hg38 assembly. For estimating the content of RNA biotypes in libraries, reads were aligned to the hg38 assembly using STAR. Subsequently, counting was performed with HTSeq-count using the geneset annotated by GENCODE Release 21 (GRCh38) and using the GTF feature 'gene' for counting.

### Proteomic analysis of the enriched RBPs

After releasing the PP enriched RBPs by RNase A elution, the RBPs were digested by filter aided sample preparation method (FASP) (29). Stable isotopic dimethyl labelling was conducted as previously described (30) for quantitative comparison between the enrichment group of the RBPs and the control. Detailed experimental procedures were provided in the Supporting information.

**Proteomic identification by LC-MS/MS.** The FASP digested and dimethyl labelled peptide samples were dissolved in 0.1% FA and were loaded into an in-house made 15 cm length reverse phase column (150 nm id) packed with Ultimate XB-C<sub>18</sub> 1.9  $\mu$ m resin (Welch materials). An Easy nLC 1000 system (Thermo) was used to separate the peptides using the following gradient: 5–8% B for 8 min, 8–22% B for 50 min, 22–32% B for 12 min, 32–90% B for 1 min, and 90% B for 7 min (A is 0.1% formic acid in water and B is 0.1% formic acid in acetonitrile). A constant flow rate of 600 nl/min was applied. The eluted peptides were sprayed into an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo) equipped with a nano-electrospray ionization source. The mass spectrometer was operated in data-dependent mode with a full MS scan (300–1400  $m/z$ ) at a resolution of 120 000, a maximum injection time of 100 ms and an AGC target value of 5e5, followed by Higher-energy Collision Dissociation (HCD) with 32% normalized collision energy. The MS2 spectra were acquired in the ion trap with an AGC target value of 5000 and a maximum injection time of 35 ms. The dynamic exclusion was set to 18 s.

**Mass spectrometric data analysis.** The raw data files were searched using MaxQuant (version 1.5.2.8) against UniProt database (release on 2015, 20 198 entry). Trypsin was set as the digestion enzyme with a maximum of two missed cleavages and the minimal peptide length was set to six amino acids. Carbamidomethyl cysteine was set as fixed modification and methionine oxidation and acetyl N-terminal were set as variable modifications. For peptide identification, the mass tolerances were 20 ppm for precursor ions and 0.5 Da for fragment ions. The false discovery rate was set  $\leq 1\%$  at spectra and protein level. Missing values were imputed using the minimum values in each dataset (31). For RBP filtering, proteins with a minimum of two identified unique peptides in at least two tests and a fold change of two or greater with  $P < 0.01$  in the experimental groups compared with the control groups were considered as RBPs. The abundance changes of the PP enriched RBPs at different time points after ActD treatment were acquired based on their label-free intensity provided by MaxQuant and were further

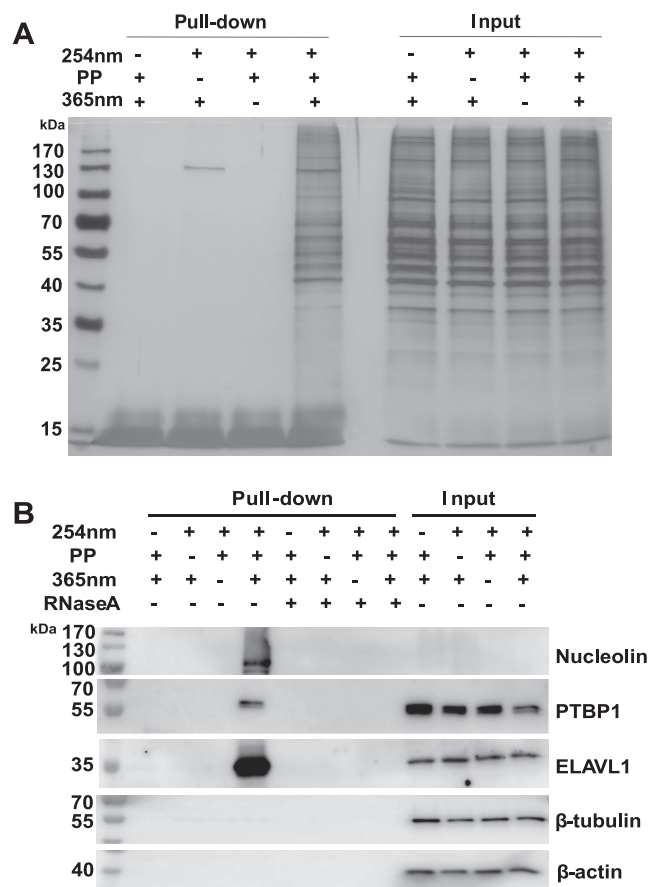
normalized by their corresponding reduction in the global proteome level to acquire the actual binding and crosslinking abundance (BCA) variation of the RBPs.

## RESULTS

### Establishment of the psoralen probe (PP)-based RNP enrichment strategy

In this work, we developed a PP-based RNA tagging strategy for the highly specific enrichment of the RNPs by exploiting the photoreactivity of psoralen analogs with uracil under 365 nm UV irradiation (Scheme 1). First, 254 nm UV irradiation was applied to living cells to establish crosslinking between RNA and RBPs. After cell lysis, PP was introduced to the cell lysate to covalently tag both coding and noncoding RNAs under 365 nm UV irradiation. The PP-tagged RNPs and RNAs were next captured by streptavidin beads and subjected to either RNA sequencing or mass spectrometry analysis. The synthetic route of the bi-functional PP is shown in Supplementary Figure S1A (Supporting Information). PP is composed of a uracil reactive psoralen, a biotin and a spacer arm. HNMR was carried out to confirm the successful synthesis of PP (Supplementary Figure S1B and C). The feasibility of PP-based RNP isolation was first demonstrated by gel electrophoresis and RNA staining (Supplementary Figure S2). RNA bands are discovered after treating the isolation products with proteinase K to remove proteins from RNAs. In contrast, the RNA bands are completely abolished upon RNase A digestion or by the omission of PP, indicating the binding and isolation RNPs by PP.

The PP-enriched RNPs were further characterized by SDS-PAGE after eluting the beads-captured RBPs by RNase A treatment. Clear protein bands of the PP enrichment products are displayed in Figure 1A. In contrast, protein bands can barely be observed without PP or 254/365 nm UV treatment, indicating a high selectivity of this enrichment method with only marginal nonspecific adsorption of non-RBPs. Using RNase A containing buffer in the washing step results in the disassociation of the RNPs and the loss of the RBPs before elution, therefore serving as a stringent negative control for this method. As expected, the protein bands are almost completely abolished by RNase A washing before elution, confirming the RNA dependence of the observed protein bands (Supplementary Figure S3). The successful enrichment of the RBPs was further demonstrated by the detection of known RBPs in western-blotting analysis, as shown in Figure 1B. ELAVL1, nucleolin and PTBP1 were clearly enriched by PP and were undetectable in the negative controls without PP or 254/365 nm UV treatment. Similar to the large-scale enrichment, RNase A treatment depletes the protein bands of three RBPs, confirming that these RBPs were directly bound and crosslinked to RNAs. The high selectivity of this method was also proven by  $\beta$ -tubulin and  $\beta$ -actin. No obvious sign of non-specific adsorption of these two non-RBPs was found in the PP enrichment products. Next, the concentration of PP and the reaction time with cell lysates were evaluated. As shown in Supplementary Figure S4A and B, optimized enrichment was obtained with 5  $\mu$ M PP and 4 min 365 nm UV

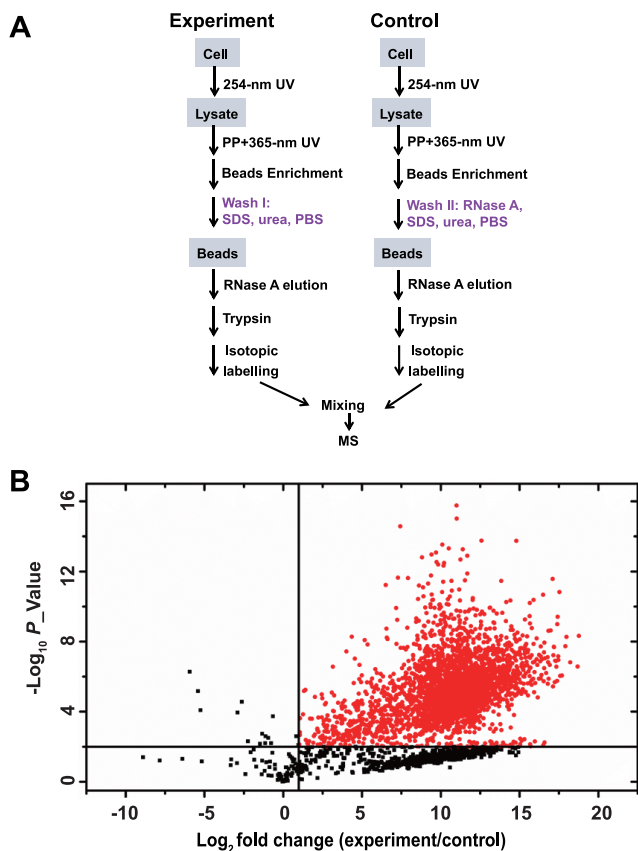


**Figure 1.** (A) SDS-PAGE characterization of PP enriched RBPs. (B) Western-blotting analysis of PP enriched ELAVL1, Nucleolin and PTBP1,  $\beta$ -tubulin and  $\beta$ -actin.

irradiation. No obvious degradation of RNA was observed under these condition (Supplementary Figure S4C).

### RNA sequencing and proteome identification of the RNPs obtained by PP enrichment

To further explore the type and relative distribution of the PP isolated RNAs, the captured RNPs were treated with proteinase K to digest and remove the RBPs. Control experiments were also conducted by extracting RNAs using the traditional TRIzol method. The resulting RNAs were analyzed by RNA sequencing (Supplementary Figure S5). The PP enrichment products and the TRIzol controls exhibit consistent RNA distribution patterns. In addition to rRNA, other major types of RNA were also found by PP enrichment, including protein-coding mRNAs and various types of noncoding RNAs, such as lincRNA, antisense RNA, snRNA and miRNA. These results demonstrate the comprehensiveness of PP tagging and its capability to enrich RBPs bound to various kinds of RNAs, although psoralen analogs were originally used for probing double-stranded regions of RNAs (23). We noticed the relatively high content of mRNA in the PP enrichment product, which may be attributed to underestimation of tRNA, since it is small and highly modified and difficult to sequence by standard methods (13,32). Presence of tRNA was revealed using gel



**Figure 2.** (A) Experimental design of the quantitative differential proteomic comparison between the experiment group and control group for RBPs identification. (B) Scatter plot displaying the  $\log_2$  fold change (x-axis) and  $-\log P$  values (y-axis) for RBPs identification by quantitative differential proteomic comparison between the experiment group and control group.

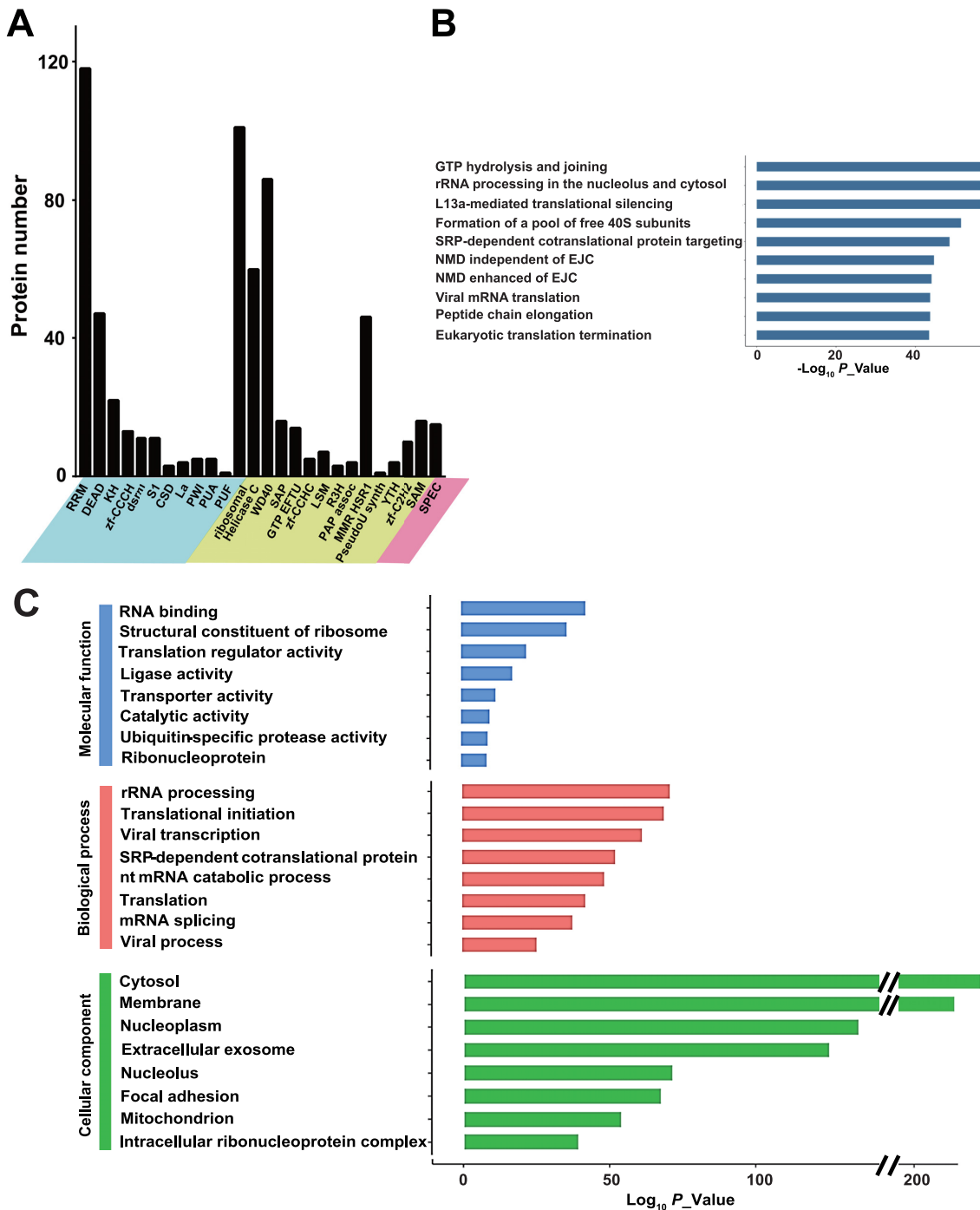
electrophoresis by the peak around 100 nt in Supplementary Figure S6.

For large-scale RBP profiling by mass spectrometry, it is crucial to differentiate and remove the non-RBPs in the results. To evaluate the proper and stringent control conditions for obtaining reliable RBPs, quantitative proteomics analysis was carried out using the Experiment and the Control design illustrated in Supplementary Table S1. As shown in Supplementary Figure S7, Control 4 showed stronger intensity than the other three controls, indicating it included more false positive hits. This result was not unexpected, since Control 4 had gone through all the treatments that may introduce false positive hits (254 nm UV irradiation, PP-treatment & 365 nm UV irradiation), while the other three controls had gone through only two of the three treatments. Therefore, Control 4 was chosen for further investigation of the PP-assay using the scheme shown in Figure 2A. Washing with RNase A before elution in the control group degrades RNA and removes the RNPs from the sample. In contrast, RNase A is not applied in the washing step, but in the elution step for the experiment group, which leads to specific elution of the beads-PP captured RBPs without the possible co-elution of DNA-binding proteins. Furthermore, a stringent screening cut-off was applied, which re-

quired a fold change  $>2$  and a  $P$  value  $<0.01$  with a minimum of two identified unique peptides in at least two tests. In this way, we identified 2986 highly confident RBPs with good reproducibility (Pearson correlations  $> 0.85$  in three technical replicates) (Figure 2B, Supplementary Figure S8 and Table S2 of the Supporting Information).

Domain analysis of the identified RBPs shows that most of the top represented domains comprise classical RBDs, such as RRM, KH and DEAD domains as well as other non-classical RBDs (Figure 3A). Interestingly, we found ‘SPEC’ domain was also enriched in our RBPs, which was not previously reported as a RBD. We conducted nucleotide-crosslinked peptide analysis of the enriched RBPs to locate the RNA-binding site and to further determine the identity of ‘SPEC’ using reported methods (15). MS analysis of the enriched cyclic-U modified peptides resulted in seven RBPs identified with SPEC domain (Supplementary Table S4). The high RNA correlation of the identified RBPs is confirmed by the top enriched pathways, which are almost all associated with RNA processing, such as ribosomal subunit formation and hydrolysis, rRNA processing, nonsense mediated decay and mRNA translation (Figure 3B). Consistently, gene ontology (GO) analysis in Figure 3C reveals highly enriched GO terms related to RNA. The most enriched molecular function and biological process terms exhibit an over-representation of RNA binding and other RNA related GO terms, suggesting that the dataset contains a large number of known RBPs. Other GO terms, such as ‘Catalytic activity’ and ‘Ubiquitin-specific protease activity’ are also enriched in the molecular function. In the ‘catalytic activity’ category, many nucleotide-binding and ATP-binding proteins were found, which is consistent with previous studies (13,14). For ‘ubiquitin-specific protease activity’, more than twenty E3 ubiquitin ligases were previously found overlap with RNA-binding proteome (33,34) and we further expanded this category by identifying  $>50$  E3 ubiquitin ligases/transferases and related proteins. Interestingly, extracellular exosome was found to be the fourth top enriched term in the cellular component analysis. Further comparison with the exosome protein database (ExoCarta) showed that 83 of the top 100 most identified exosome proteins were found in our RBPs. Such high coverage of exosome proteins was not reported in other studies using either polyA-dependent or polyA-independent enrichment strategies. Considering that RNAs in exosomes have been extensively studied in recent years as functional regulators and biomarkers for disease diagnostics and that few reports on their interacting proteome have been published, the PP-based strategy may facilitate the deeper exploration of RBPs in extracellular exosome.

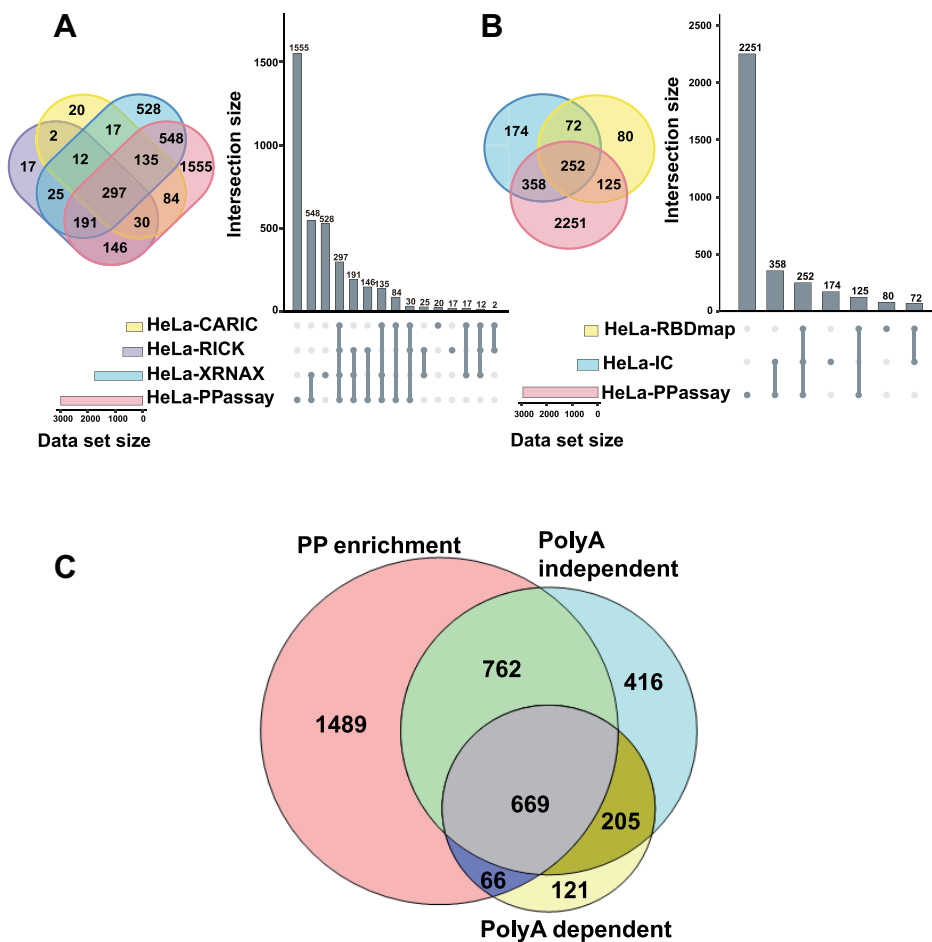
Compared with the reported RBPs enriched from HeLa cells by polyA-dependent and polyA-independent strategies (11,13–15,28), the products of PP enrichment show overlapping but distinct results. As displayed in Figure 4, PP-based enrichment not only covered majority of the reported RBPs in HeLa cells, but also provided  $>1000$  additional RBPs, presumably due to the facile and efficient RNA tagging via the cycloaddition reaction by the psoralen group of PP, which is capable of pulling down both coding and non-coding RBPs. We noticed that even after combining three large-scale RBP data obtained by polyA independent



**Figure 3.** (A) Number of identified RBPs with classical (left) and non-classical RBDs (right). (B) pathway and (C). GO enrichment analysis of the obtained RBPs.

methods and ours, 121 RBPs were still exclusively identified by the polyA dependent enrichment methods (Figure 4C). Failure to cover these RBPs by the polyA independent methods may be attributed to the low abundance of coding RNA in the total RNA. Coding RNA binding proteins may be overwhelmed by the more abundant non-coding RNA binding proteins and lost in mass spectrometry analysis. Next, we compiled an MS identified RBP dataset of 4758 RBPs from studies using different cell lines. Although the

size of RBP data has been substantially expanded by the recently reported RBP enrichment methods, our PP-based enrichment still provided 782 candidate RBPs that were not found in previous large-scale RBP profiling (Supporting Information Supplementary Table S3). We compared the physicochemical characteristics of the PP identified all RBPs, 782 candidate RBPs and that of the literature reported RBPs, including positively charged amino acid, isoelectric point, hydrophobicity and disorder. The RNA-binding



**Figure 4.** UpSet plots comparing the RBPs identified by PP-assay and other polyA independent (A) and dependent (B) RBP enrichment methods. (C) Venn diagram showing overlap of the RBPs identified by PP, reported polyA dependent and independent enrichment methods.

proteomes were divided into polyA interactome and non-polyA interactome following the way described previously (15). As shown in Supplementary Figure S9, the distribution curves of both of the PP identified all RBPs and candidate RBPs resemble that of the previously known RBPs and fall within the range of those of the RBPs obtained by phase separation-based methods (OOPS and XRNAX) (15,16) or metabolic labelling-based methods (RICK and CARIC) (13,14). Interestingly, the non-poly(A)-binding proteins are clearly distinct when compared to poly(A)-binding proteins. The poly(A)-binding RBPs are less hydrophobic and obviously more disordered. Whereas the non-poly(A)-binding RBPs do not have bimodal isoelectric points (IPs) distribution and generally have more acidic IPs and less content of positively charged amino acids. The 782 candidate RBPs (red) exhibit an obviously lower abundance distribution in Supplementary Figure S10A, while the previously reported RBPs (blue) found in this work are more concentrated in region of higher abundance. Similar trends are also found in Supplementary Figure S10B. The median abundance of the candidate RBPs is ~10-fold lower than that of the reported RBPs. To further explore the function of these candidate RBPs, we conducted GO enrichment analysis of the 782 candidate RBPs that were not reported by other

works (11–16,35,36) (Supplementary Figure S11A). Interestingly, we found that the GO terms ‘membrane’ and ‘GT-Pases activity’ are among the top enriched GO terms in the cellular component and molecular function. Although not previously reported as RBPs, proteins in both terms regulate protein transport and membrane trafficking via binding with RNA for correct targeting towards specific organelles (37,38). DNA-binding RBPs were previously reported via enrichment using the serIC method (39). 75% of the DNA-binding RBPs enriched by serIC were also identified our PP-assay (Supplementary Table S4). Further inspection on the 782 candidate RBPs revealed additional 55 DNA-binding proteins (DBPs) not discovered in previous work, including [poly [ADP-ribose] polymerase 2 (PARP-2), that are involved in diversified functions. Although the RNA-binding abilities of these DBPs have not been discovered in previous large-scale RBP identification, the SAP domain containing PARP-2 was reported to bind to specific transcripts for RNA regulation in an individual protein study (40), indicating the reliability of these candidate RBPs. Furthermore, GO (biological process) and pathway analysis in Supplementary Figure S11A both exhibited overrepresented ‘Metabolism’ among the candidate RBPs, especially the citric acid (TCA) cycle. TCA cycle have long been related to

RNA-binding activities. PP enrichment approach managed to identify 70 RBPs that associated with TCA cycle and covered more than 80% of the previously reported TCA cycle-related RBPs (4,5) (Supplementary Figure S11B). The binding between RNA and numerous metabolic enzymes also has been previously documented to regulate the expression of the bound mRNA in response to metabolite levels or, conversely, to control the activity of the bound enzyme (41). Compared with the RBPs identified by PP-assay, 79 of the 104 reported RNA-binding enzymes (42,43) were covered (Supplementary Table S4). Based on the Reactome pathway analysis (44), we found that the 782 candidate RBPs include 178 metabolism related proteins (Supplementary Figure S11C) that are distributed among a variety of pathways, including TCA cycle, carbohydrate and lipid metabolism. These candidates expanded the scale of RBP, suggesting that the interplay between RNA, DNA, membrane, protein and metabolite may be broader than previously realized and deserves further exploration.

Next, we selected WDR82, ARL2, ARF5 and ELP3 from the 782 candidate RBPs that were not reported in previous large-scale RBP investigations for validation via cross-linking and immunoprecipitation (CLIP) and western blotting (45,46). The four proteins were selected from different function categories and subcellular locations to avoid bias towards a particular subgroup. WDR82 is a chromatin binding protein and mainly locates in nucleus (47). ARL2 regulates formation of new microtubules and centrosome integrity and locates in cytoplasm, cytoskeleton, mitochondrion and nucleus (48). ARF5 is a GTP-binding protein involved in protein trafficking and may modulate vesicle budding, and locates in cytoplasm, Golgi apparatus and membrane (49). ELP3 is a RNA polymerase II core binding protein involved in transcriptional elongation and locates in cytoplasm and nucleus (50). FLAG-tagged candidate RBPs were expressed in HeLa cells and irradiated with 254-nm UV light. The tagged RBPs were pulled down together with the cross-linked RNA and characterized by western blotting, after shortening the RNA and labelling it with biotin. As shown in Supplementary Figure S12A, similar to the positive control using known RBP (ELAVL1), the positive existence of RNA was revealed by the corresponding gel bands for the four candidate RBPs and demonstrated their RNA-binding activity. To further evaluate our validation result, we performed PP isolation assay to pull-down all the RNPs and characterized the four candidate RBPs (ELP3, WDR82, ARF5 and ARL2) by antibody-based western blotting. As shown in Supplementary Figure S12B, corresponding protein bands were found only in the experiment group (with 254, 365 nm UV-irradiation, PP treatment, but not RNase A washing) and the input group. No false positive bands were discovered in any of the control groups.

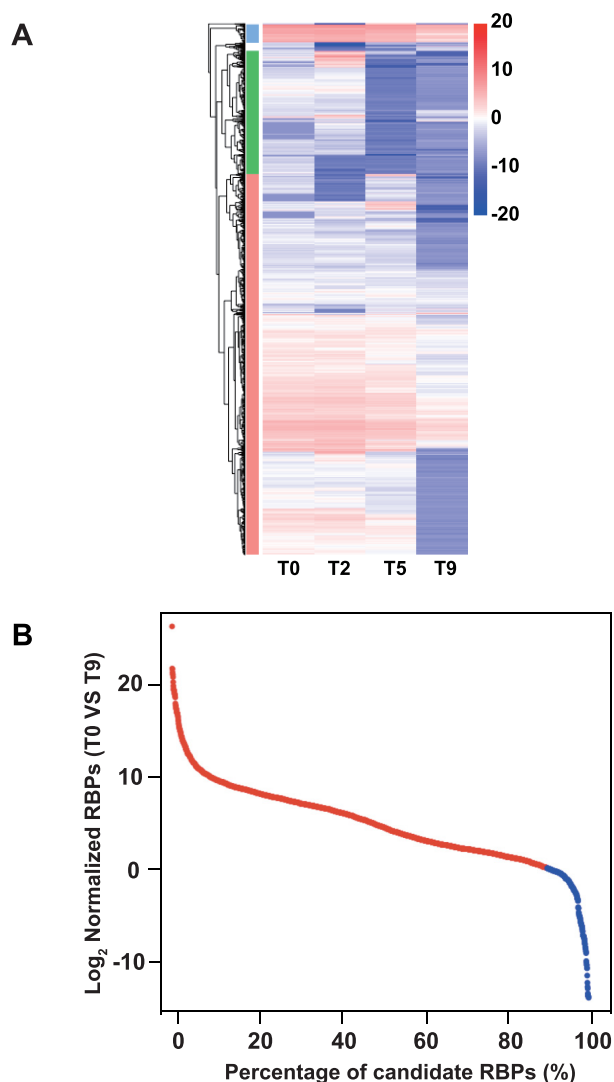
#### **Actinomycin D (ActD)-assisted RNP stability/dynamics study and large-scale evaluation of the candidate RBPs**

RBPs regulate various biological processes by associating and disassociating with RNAs, therefore the stability and dynamics of the corresponding RNPs under different cellular states hold great significance. We conducted a large-

scale investigation on the stability/dynamics of RNPs under ActD treatment using our PP-based enrichment strategy. ActD blocks RNA transcription and leads to decreasing levels of RNAs and disassociation of the RNPs in the cell. As shown in Supplementary Figure S13, the levels of various kinds of RNAs in the cell continually decrease after ActD treatment, with protein-coding RNA displaying the most distinct drop of ~60%. Similarly, the inhibition of RNA transcription results in a decreased number of PP-enriched RBPs, with 2137, 1921, 1742 and 786 RBPs (Supplementary Table S2) identified at 0, 2, 5 and 9 h after ActD treatment, respectively, indicating increased disassembly of RNPs due to the reduced RNA levels. The dynamics of RBPs in the RNPs were determined by quantitative comparison of the PP-enriched RBPs in the ActD-treated cells with the corresponding RBPs in the untreated cells. The quantity of the RBPs obtained by PP-assay is controlled by UV crosslinking between RNA and protein, which may be influenced by multiple factors, such as binding geometry of RNA, nucleotide and amino acids composition, and duration of the interaction (51). Therefore, the same UV-irradiation condition must be applied for the cells treated with different ActD conditions to minimize the variation induced by UV crosslinking. Furthermore, to exclude the reduced abundance of RBPs caused by protein degradation and RNA synthesis inhibition, the global proteome level of each RBP at different time points during ActD treatment was determined by proteome profiling. After normalizing the change in abundance of the PP-enriched RBPs by the corresponding reduction in their global proteome levels to obtain their actual binding and crosslinking abundance (BCA) variation, the true dynamic trends of RBPs in the RNP are shown in Figure 5A. Since our PP strategy actually isolates the RNPs, the dynamic trends of each identified RBP can be used as an indication of the stability of the RNPs. The stability/abundance of RNPs can be divided into three categories. Approximately 5% of the RBPs are in relatively stable RNPs, which are present in almost the same abundance throughout the ActD treatment (blue zone). An obviously greater descending trend is displayed in the green zone, with most of the RBPs decreasing below the detection limit of the PP enrichment method after 5 h RNA synthesis inhibition, indicating that these RBPs are present in unstable RNPs with high decay rates. The rest of the RBPs are in RNPs with median decay rates, which disassociate from two to nine hours after ActD treatment.

Although RNA-binding of the candidate RBPs can be validated individually by the traditional CLIP and western blotting strategy as mentioned above, the particularly low-throughput of this method is a severe bottleneck for the large-scale evaluation of the hundreds to thousands of candidate RBPs discovered in recent years. Inspired by the ActD induced inhibition of RNA transcription and the dynamic changes of RNPs, we propose a high-throughput way to further evaluate RNA-binding of the 782 candidate RBPs that were not discovered by other large-scale RBP identification studies. As shown in Figure 5B, close to 90% (red) of the candidate RBPs display normalized BCA >0 after nine hours of ActD treatment, which may be attributed to either the decrease of the available RNA and disassembly of RNPs after inhibition of RNA synthesis or changes





**Figure 5.** (A) heat map and (B) scatter plot of normalized BCA variation of the RBPs after treating cells with ActD for 0, 2, 5 and 9 h (T0–T9). Normalized BCA of RBPs (T0 versus T9) = fold change (T0 versus T9) of RBPs in PP-assay/fold change (T0 versus T9) of the corresponding RBPs in the whole cell lysate.

in the binding activity of the RBPs, though cannot be assessed with the current available data.

## DISCUSSION

In this work, we developed a photo-induced RNA tagging and RNP enrichment strategy. We demonstrated the capability of PP tagging to enrich RBPs bound to various kinds of RNAs, although psoralen analogs were originally used for probing the double-stranded regions of RNAs (23). Using this strategy, we successfully identified 2986 RBPs from HeLa cells, including 782 candidate RBPs that were not discovered in previous large-scale studies. PP strategy is simple and does not need metabolic labelling. The high enrichment specificity and free of possible contamination of glycoproteins (16) make this approach advantageous in dis-

covering new candidate RBPs. Interestingly, GO analysis of the 782 candidate RBPs reveals that metabolism and related terms are overrepresented in molecular function and biological process. The 178 metabolism pathways related proteins in the candidate RBPs indicate a possible deeper involvement of RBPs in the regulation of metabolic process than previously known, and the mechanisms of this regulation are yet to be investigated. Furthermore, PP can be used as a powerful tool to study the dynamic association of RNPs in the cell. Large-scale investigation on the perturbations of RNPs during RNA transcription inhibition using our PP-based enrichment strategy reveals the first distribution patterns of RBPs bound to RNA with different decay rates. More importantly, this strategy provides a high throughput way to evaluate candidate RBPs by exploiting their obviously greater abundance decrease in the RNPs than that in the global proteome due to the continued dissociation of the RNPs after the inhibition of RNA synthesis. Though successfully applied in a non-polyA dependent manner and is theoretically capable of enriching both coding and non-coding RNA binding-proteins at the same time, there are still some limitation in the PP-assay. Considering coding RNAs are only 3% in the total RNAs, many of the coding RNA-binding-proteins may be overwhelmed by the more abundant non-coding RNA-binding proteins in the enrichment products and leads to biased RBPs identification. Therefore, for studies targeting at mRNA-binding proteins, polyA dependent enrichment methods should be a better choice (51).

## DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023162. All RNA-seq data used in this manuscript have been deposited in Gene Expression Omnibus ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)). The data about the type and relative distribution of the PP isolated RNAs were deposited under accession number GSE163230. The data about the dynamic investigation of the RNA upon transcription inhibition by ActD treatment were deposited under accession number GSE151805. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Key R&D Program of China [2016YFA0501403, 2018YFC0910302, 2017YFA0505002]; National Key Laboratory of Proteomics [SKLP-K201706]; National Natural Science Foundation of China [32088101, 21675172]. Funding for open access charge: National Key R&D Program of China [2016YFA0501403].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Müller-McNicoll, M. and Neugebauer, K.M. (2013) How cells get the message: dynamic assembly and function of mRNA–protein complexes. *Nat. Rev. Genet.*, **14**, 275–287.
2. Hocq, R., Paternina, J., Alasseur, Q., Genovesio, A. and Le Hir, H. (2018) Monitored eCLIP: high accuracy mapping of RNA–protein interactions. *Nucleic Acids Res.*, **46**, 11553–11565.
3. Hudson, W.H. and Ortlund, E.A. (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.*, **15**, 749–760.
4. Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.
5. Liao, Y., Castello, A., Fischer, B., Leicht, S., Föehr, S., Frese, C.K., Ragan, C., Kurscheid, S., Pagler, E. and Yang, H. (2016) The cardiomyocyte RNA-binding proteome: links to intermediary metabolism and heart disease. *Cell Rep.*, **16**, 1456–1469.
6. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. and Krzyzosiak, W.J. (2016) Identifying proteins that bind to specific RNAs-focus on simple repeat expansion diseases. *Nucleic Acids Res.*, **44**, 9050–9070.
7. Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A.I., González-Almela, E., Lenz, C.E., Bach-Pages, M., Cox, V., Avolio, R. and Davis, T. (2019) System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell*, **74**, 196–211.
8. Ramanathan, M., Porter, D.F. and Khavari, P.A. (2019) Methods to study RNA–protein interactions. *Nat. Methods*, **16**, 225–234.
9. 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.
10. Lee, F.C. and Ule, J. (2018) Advances in CLIP technologies for studies of protein–RNA interactions. *Mol. Cell*, **69**, 354–369.
11. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T. and Steinmetz, L.M. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
12. Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K. and Milek, M. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
13. Huang, R., Han, M., Meng, L. and Chen, X. (2018) Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3879–E3887.
14. Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y. and Pulido-Quetglas, C. (2018) Capturing the interactome of newly transcribed RNA. *Nat. Methods*, **15**, 213–220.
15. 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.
16. Queiroz, R.M., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F. and Dezi, V. (2019) Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.*, **37**, 169–178.
17. Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S. and Selbach, M. (2019) Purification of cross-linked RNA–protein complexes by phenol-toluol extraction. *Nat. Commun.*, **10**, 990.
18. Asencio, C., Chatterjee, A. and Hentze, M.W. (2018) Silica-based solid-phase extraction of cross-linked nucleic acid–bound proteins. *Life Sci Alliance.*, **1**, e201800088.
19. Shchepachev, V., Bresson, S., Spanos, C., Petfalski, E., Fischer, L., Rappsilber, J. and Tollervy, D. (2019) Defining the RNA interactome by total RNA-associated protein purification. *Mol. Syst. Biol.*, **15**, e8689.
20. Smith, T., Villanueva, E., Queiroz, R.M., Dawson, C.S., Elzek, M., Urdaneta, E.C., Willis, A.E., Beckmann, B.M., Krijgsveld, J. and Lilley, K.S. (2020) Organic phase separation opens up new opportunities to interrogate the RNA-binding proteome. *Curr. Opin. Chem. Biol.*, **54**, 70–75.
21. Periat, A., Krull, I.S. and Guillaume, D. (2015) Applications of hydrophilic interaction chromatography to amino acids, peptides, and proteins. *J. Sep. Sci.*, **38**, 357–367.
22. Zhu, R., Zacharias, L., Wooding, K.M., Peng, W. and Mechref, Y. (2017) Glycoprotein enrichment analytical techniques: advantages and disadvantages. *Methods Enzymol.*, **585**, 397–429.
23. Calvet, J.P. and Pederson, T. (1981) Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking in vivo. *Cell*, **26**, 363–370.
24. Garrett-Wheeler, E., Lockard, R.E. and Kumar, A. (1984) Mapping of psoralen cross-linked nucleotides in RNA. *Nucleic Acids Res.*, **12**, 3405–3424.
25. Thompson, J.F., Bachelier, J.P., Hall, K. and Hearst, J.E. (1982) Dependence of 4′-(hydroxymethyl)-4, 5′, 8-trimethylpsoralen photoaddition on the conformation of RNA. *Biochemistry*, **21**, 1363–1368.
26. Reim, A., Ackermann, R., Font-Mateu, J., Kammel, R., Beato, M., Nolte, S., Mann, M., Russmann, C. and Wierer, M. (2020) Atomic-resolution mapping of transcription factor–DNA interactions by femtosecond laser crosslinking and mass spectrometry. *Nat. Commun.*, **11**, 3019.
27. Steube, A., Schenk, T., Tretyakov, A. and Saluz, H.P. (2017) High-intensity UV laser ChIP-seq for the study of protein–DNA interactions in living cells. *Nat. Commun.*, **8**, 1303.
28. Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Föehr, S., Curk, T., Krijgsveld, J. and Hentze, M.W. (2016) Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell*, **63**, 696–710.
29. Wiśniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods*, **6**, 359–362.
30. Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. and Heck, A.J. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.*, **4**, 484–494.
31. Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L. and Hu, B. (2019) Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*, **567**, 257–261.
32. Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M. and Pan, T. (2015) Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods*, **12**, 835–837.
33. Kwon, S.C., Yi, H., Eichelbaum, K., Föehr, 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.
34. Thapa, P., Shanmugam, N. and Pokrzywa, W. (2020) Ubiquitin signaling regulates RNA biogenesis, processing, and metabolism. *Bioessays*, **42**, 1900171.
35. Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A. and Schwinn, M.K. (2016) SONAR discovers RNA-binding proteins from analysis of large-scale protein–protein interactomes. *Mol. Cell*, **64**, 282–293.
36. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 829–845.
37. Béthune, J., Jansen, R.-P., Feldbrügge, M. and Zarnack, K. (2019) Membrane-associated RNA-binding proteins orchestrate organelle-coupled translation. *Trends Cell Biol.*, **29**, 178–188.
38. Clement, N. and Polacek, N. (2010) Ribosome-associated GTPases: the role of RNA for GTPase activation. *RNA Biol.*, **7**, 521–527.
39. Conrad, T., Albrecht, A.-S., de Melo Costa, V.R., Sauer, S., Meierhofer, D. and Ørom, U.A. (2016) Serial interactome capture of the human cell nucleus. *Nat. Commun.*, **7**, 11212.
40. Bock, F.J., Todorova, T.T. and Chang, P. (2015) RNA regulation by poly (ADP-ribose) polymerases. *Mol. Cell*, **58**, 959–969.
41. Castello, A., Hentze, M.W. and Preiss, T. (2015) Metabolic enzymes enjoying new partnerships as RNA-binding proteins. *Trends Endocrinol. Metab.*, **26**, 746–757.
42. Matia-González, A.M., Laing, E.E. and Gerber, A.P. (2015) Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat. Struct. Mol. Biol.*, **22**, 1027–1033.

43. Beckmann,B.M., Horos,R., Fischer,B., Castello,A., Eichelbaum,K., Alleaume,A.-M., Schwarzl,T., Curk,T., Foehr,S. and Huber,W. (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.*, **6**, 10127.
44. Fabregat,A., Jupe,S., Matthews,L., Sidiropoulos,K., Gillespie,M., Garapati,P., Haw,R., Jassal,B., Korninger,F. and May,B. (2018) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.
45. Chen,T., Hao,Y.-J., Zhang,Y., Li,M.-M., Wang,M., Han,W., Wu,Y., Lv,Y., Hao,J. and Wang,L. (2015) m6A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell stem cell*, **16**, 289–301.
46. Ping,X.-L., Sun,B.-F., Wang,L., Xiao,W., Yang,X., Wang,W.-J., Adhikari,S., Shi,Y., Lv,Y. and Chen,Y.-S. (2014) Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.*, **24**, 177–189.
47. Lee,J.-H. and Skalnik,D.G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.*, **28**, 609–618.
48. Bhamidipati,A., Lewis,S.A. and Cowan,N.J. (2000) ADP ribosylation factor-like protein 2 (Arl2) regulates the interaction of tubulin-folding cofactor D with native tubulin. *J. Cell Biol.*, **149**, 1087–1096.
49. Egerer,J., Emmerich,D., Fischer-Zirnsak,B., Chan,W.L., Meierhofer,D., Tuysuz,B., Marschner,K., Sauer,S., Barr,F.A. and Mundlos,S. (2015) GORAB missense mutations disrupt RAB6 and ARF5 binding and golgi targeting. *J. Invest. Dermatol.*, **135**, 2368–2376.
50. Kim,J.-H., Lane,W.S. and Reinberg,D. (2002) Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 1241–1246.
51. Perez-Perri,J.I., Noerenberg,M., Kamel,W., Lenz,C.E., Mohammed,S., Hentze,M.W. and Castello,A. (2020) Global analysis of RNA-binding protein dynamics by comparative and enhanced RNA interactome capture. *Nat. Protoc.*, **16**, 27–60.