Nucleic Acids Research, 2026, **54**, gkag090
https://doi.org/10.1093/nar/gkag090
Methods

RAPseq enables large-scale identification of RBP–RNA interactions and reveals essentials of post-transcriptional gene regulation

Riccardo Mosca[†], **Carlos J. Gallardo-Dodd**[†], **Qun Li**[†], **Christian Sommerauer**, **Justas Šidiškis**, **Jonas Nørskov Søndergaard**, **Claudia Kutter**[*]

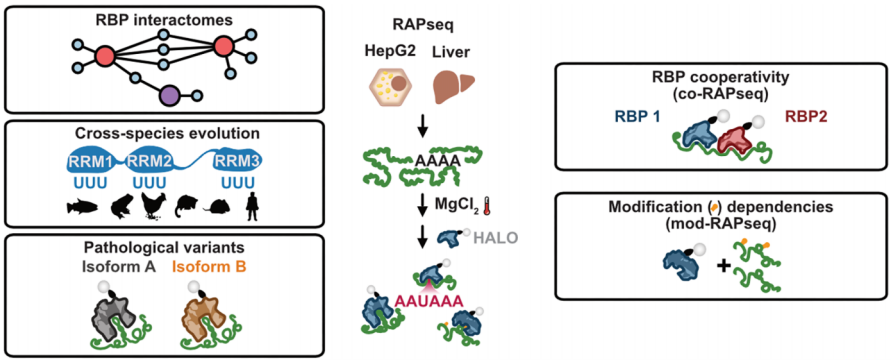
Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Science for Life Laboratory, 171 11 Stockholm, Sweden

[*]To whom correspondence should be addressed. Tel: +46 76 493 38 96; Fax: +46 8 31 11 01; Email: claudia.kutter@ki.se
[†]Riccardo Mosca, Carlos J. Gallardo-Dodd, and Qun Li contributed equally to this work.

Abstract

Over the past decade, thousands of putative human RNA-binding proteins (RBPs) have been identified, increasing the need for methods that define their RNA-binding capacities across diverse biological settings. Existing methods rely either on antibody-based *in vivo* capture (e.g. CLIP-seq), which depends on cross-linking efficiency and antibody availability, or on synthetic oligonucleotide-based assays (e.g. RNAcompete), which use artificial RNA substrates and cannot assess binding across the native transcriptome. To bridge this gap, we developed RNA affinity purification followed by sequencing (RAPseq), an *in vitro* method that profiles RBP-binding to native cellular RNA, enabling large-scale transcriptome-wide characterization of RNA–protein interactions without antibodies or synthetic probes. Using RAPseq, we characterized the RNA interactomes of 11 canonical RBPs and 26 non-canonical RBPs, and uncovered novel and specialized moonlighting RNA-binding activities. Applying RAPseq to vertebrate HUR proteins revealed recognition of a conserved RNA-binding motif but showed species-specific binding preferences. Profiling of five pathological IGF2BP family variants exhibited distinct gain- and loss-of-function binding patterns, with implications for cancer biology. Our combinatorial RBP-binding assay (co-RAPseq) uncovered cooperative RNA-binding by HUR and PTBP1, including *de novo* estimation of the optimal binding distance. Lastly, we introduce a modification-sensitive assay (mod-RAPseq) to distinguish between modification-dependent and -independent RNA-binding sites of YTHDF1 and YBX1. Overall, our simple, scalable, and versatile method enables exploration of complex RNA–protein interactions and the regulatory layers that shape post-transcriptional gene regulation.

Graphical abstract



Introduction

RNA-binding proteins (RBPs) determine the fate and function of all RNA species from synthesis to decay. The latest advances in cross-linking-based methods coupled with mass spectrometry have enabled the detection of RBP–RNA complexes and have extended the number of putative human RBPs to approximately 4200 [1–4]. These proteins can be grouped into canonical RBPs (cRBPs) or non-canonical RBPs (ncRBPs) depending on the presence or absence, respectively, of a biochemically characterized RNA-binding domain. Parallel developments in mapping cellular RNA modifications transcriptome-wide have revealed an additional layer of complexity and uncovered extensive changes that impact RBP-binding [5–10]. Modifications, such as N^6-methyladenosine (m^6A) and 5-methylcytosine (m^5C), are now recognized as key regulatory layers in post-transcriptional control [6, 11, 12].

Received: June 27, 2025. **Revised:** December 18, 2025. **Accepted:** January 19, 2026

To dissect mechanisms of RBP-mediated regulatory processes, it is crucial to identify the RNA targets of RBPs, their exact binding sites, and their interaction strengths. UV cross-linking followed by immunoprecipitation (CLIP) [13] has revolutionized the *in vivo* mapping of RBP-binding sites and coupling it to high-throughput RNA sequencing (RNAseq) has yielded many variations of the method, including HITS-CLIP [14], PAR-CLIP [15], iCLIP [16], irCLIP [17], eCLIP [18], and multiplexed eCLIP [19], among others. Although CLIPseq methods have enabled breakthroughs in the field, limitations remain. The inherent low efficiency of UV cross-linking biases the capture of distinct amino acid–ribonucleotide interactions. Furthermore, the dependence on high-quality antibodies and extensive hands-on time to produce CLIPseq libraries contribute to quantitative variability and high experimental failure rates [20, 21]. In addition, large-scale *in vitro* technologies, such as RNAcompete [22] and RNA Bind-n-Seq (RBNS) [23], were developed to profile binding of purified recombinant RBPs to random synthetic RNAs of 20 or 40 nucleotides (nt) and allow quantitative assessment of primary and secondary binding motifs. However, these methods cannot infer the impact of intrinsic RNA features present in native cellular transcripts.

To overcome current limitations in studying RNA–protein interactions, we developed RAPseq. RAPseq is a RBP-centric affinity purification-based method that allows the study of *in vitro* RBP–RNA interactomes using native cellular transcriptomes in a scalable manner. Briefly, in RAPseq, native cellular RNA is chemically fragmented to generate the binding substrate, a step that can be performed in parallel with the cloning and *in vitro* translation of Halo-tagged RBPs. The binding assay is then performed between the recombinant RBP(s) and the fragmented RNA. RBP–RNA complexes are affinity-purified, and the associated RNA is sequenced to map transcriptome-wide binding sites. Unlike existing antibody- or synthetic oligo-based approaches, RAPseq does not require immunoprecipitation or cross-linking, preserves endogenous RNA features, including modifications, and is compatible with high-throughput, multiplexed analysis. The use of recombinant RBPs generated in cell-free systems and native RNA makes RAPseq uniquely suited for functional and comparative analyses across a wide range of RBPs.

We applied RAPseq in multiple distinct settings to demonstrate its versatility and to explore previously inaccessible questions in RNA biology. First, we systematically mapped transcriptome-wide RNA-binding activities of 11 cRBPs and 26 ncRBPs not previously studied in this context. This allowed us to uncover widespread moonlighting RNA-binding capacities. Second, we investigated the evolutionary binding propensities of the RBP HUR (ELAVL1) across vertebrates. HUR is a well-characterized post-transcriptional regulator that was selected because of its conserved role in mRNA stabilization. Cross-species RAPseq allowed us to identify commonalities in RNA motif recognition and differences in target site selection. Third, we examined disease-associated consequences of five disease-associated variants of IGF2BP paralogs, which are frequently mutated in cancer. RAPseq revealed that the pathological variants exhibit distinct RNA-binding patterns, providing insight into their roles in tumorigenesis. Fourth, we employed co-RAPseq to investigate cooperative RNA-binding between two RBPs, HUR and PTBP1, due to their similarities in target motif preferences. This enabled transcriptome-wide mapping of cooperative binding sites and *de novo* estimation of the optimal spacing between binding motifs. Finally, we developed mod-RAPseq to distinguish between modification-dependent and -independent RNA-binding events. We focused on YTHDF1 and YBX1, representative readers of m^6A- and m^5C-marked RNAs, respectively. By comparing cellular (RAPseq) and *in vitro* generated (mod-RAPseq) transcriptomes, we quantified the contribution of RNA modifications to RBP recognition.

Altogether, these implementations highlight the broad applicability and unique capabilities of RAPseq and its variants in advancing our understanding of RNA–protein interactions. Through its diverse applications, RAPseq uncovers novel conceptual and mechanistic insights into the evolution dynamics, cooperative behavior, and disease relevance of RBP–RNA-binding.

Materials and methods

RNA extraction from HepG2 cells

HepG2 cells were obtained from the American Type Culture Collection (ATCC) with a certified genotype. Cells were cultured in Dulbecco's modified Eagle's medium (Merck) supplemented with 1% penicillin/streptomycin (Merck) and 10% fetal bovine serum (HyClone), and maintained until reaching 80% confluency in 15 cm culture dishes (Sarstedt). After removing the medium, cells were washed once with phosphate-buffered saline (PBS). For RNA extraction, cells were lysed in 1 ml of QIAzol® (QIAGEN), scraped, and collected into a 2 ml microcentrifuge tube. A total of 200 µl of chloroform (Merck) was added to the lysate, followed by vigorous shaking by hand for 10 seconds and incubation at room temperature for 3 minutes. The collection tube was centrifuged at 9000 *g* at 4°C for 10 minutes. The aqueous phase was collected and mixed with an equal volume of isopropanol (Merck), shaken vigorously, and incubated at room temperature for 10 minutes. RNA was precipitated by centrifugation at 9000 *g* at 4°C for 10 minutes, washed twice with 500 µl of 80% ethanol (Avantor), and resuspended in nuclease-free water (Thermo Fisher Scientific) at a final concentration of 600 ng μl^{-1} measured by a Qubit™ 4 fluorometer (Thermo Fisher Scientific). For removal of genomic DNA, DNase treatment was performed at 37°C for 30 minutes by adding 2 U of TURBO DNase™ (Thermo Fisher Scientific) and 5 µl of 10× TURBO DNase buffer per 10 µg of RNA. RNA was then purified using RNA Clean & Concentrator™ columns (Zymo Research) according to the manufacturer's instructions.

RNA extraction from vertebrate livers

Mouse livers (2.5 months old, male) were obtained from the Cambridge Institute, UK, under Home Office license PPL 80/2197. Opossum livers (17 months old, males) were provided by the University of Glasgow, UK. Chicken livers (2 years old, female) were purchased from the Poultry Production Unit at the BBSRC Institute of Animal Health, Compton, UK. Frog livers (2 months old) were obtained from the Wellcome Trust/Cancer Research UK Gurdon Institute (Cambridge, UK). RNA was extracted as previously described [24]. Adult male zebrafish liver tissue was provided by the Zebrafish Core Facility at Comparative Medicine, Karolinska Institute, Sweden. Husbandry and breeding were conducted in accordance with the ethical permit approved by the Stockholm North Ethical Board (Dnr 14049-2019). A flash-frozen

zebrafish liver piece (approximately 20 mg) was immediately immersed in 1 ml of QIAzol® (QIAGEN) and homogenized using a stainless-steel bead-containing tube in a TissueLyser II (QIAGEN) at 25 Hz for 2 minutes, repeated three times. After complete homogenization, the bead was removed, and RNA extraction was performed as described above.

Cloning of RNA-binding proteins

HaloTag® cloning vectors for cell-free protein production, encoding either N- or C-terminal fusions, were purchased (Promega, G8031). Before cloning RBPs, the *cer* recombination sequence element was removed from the vectors by PCR amplification using primers F_remove_cer and R_remove_cer (Supplementary Table S1). PCR products were first purified with the Zymoclean Gel DNA Recovery™ kit (Zymo Research) and subsequently with the DNA Clean & Concentrator™ kit (Zymo Research), 5′-end phosphorylated with the T4 polynucleotide kinase (PNK) (New England Biolabs), and ligated using T4 DNA ligase (New England Biolabs) before transformations into One Shot™ TOP10 Chemically Competent *Escherichia coli* cells (Thermo Fisher Scientific). Plasmid was purified with ZR Plasmid Miniprep™ (Zymo Research). The resulting plasmid lacking the *cer* site was used as a backbone for all subsequent RBP cloning. RBPs were cloned with a C-terminal HaloTag, except for STMN1, which was cloned with an N-terminal HaloTag because of its short coding sequence. For RBP cDNA synthesis, 2 μg of total RNA was reverse transcribed using an oligo-dTVN primer (5′-TTTTTTTTTTTTTTTTTTVN-3′) in a 20 μl reaction with 1 μl of SuperScript™ III Reverse Transcriptase (SSIII, Thermo Fisher Scientific) at 50°C for 1 hour, according to manufacturer's instructions. By adding an extra 1 μl of SSIII, the reverse transcription reaction proceeded at 50°C for an additional hour. The reverse transcription reaction was inactivated by heating at 70°C for 20 minutes, followed by cooling to room temperature and treatment with 1 μl of RNase H (Thermo Fisher Scientific) at 37°C for 30 minutes. This protocol was applied to total RNA isolated from HepG2 cells and liver tissues from zebrafish, frog, chicken, opossum, and mouse. For each RBP, 1 μl of the reverse transcription reaction was used as a template in a 25 μl PCR using Q5® High-Fidelity DNA Polymerase (New England Biolabs) and RBP-specific forward and reverse primers (Supplementary Table S1). PCR products were purified as previously described and either directly assembled into linearized HaloTag backbones via Gibson Assembly (New England Biolabs) or restriction digested (XbaI, SbfI-HF, FspI, SnaBI, and XhoI) (New England Biolabs) and column purified prior to ligation. All RBP-Halo constructs were verified by Sanger sequencing (Eurofins). Plasmids generated in this study have been deposited with Addgene along with their sequences (Supplementary Table S2). With the exception of the frog and zebrafish *HUR* orthologs, all other orthologs were amplified using human *HUR*-specific primers (Supplementary Table S1).

Site-directed mutagenesis

Pathological IGF2BP variants were produced by site-directed mutagenesis (SDM) using variant-specific primers and wild-type clones as templates (Supplementary Table S1). SDM reactions were carried out by PCR using the Q5® polymerase (New England Biolabs).

The PCR product was then purified using the DNA Clean & Concentrator™ kit (Zymo Research), and digested with the DpnI exonuclease (New England Biolabs) to remove the methylated wild-type template. The DNA was subsequently purified and 5′-end phosphorylated using T4 PNK followed by ligation of the ends by the T4 DNA ligase (New England Biolabs) to obtain a circularized plasmid. The plasmids were then used to transform into One Shot™ TOP10 Chemically Competent *E. coli* cells (Thermo Fisher Scientific), and colonies were screened for the presence of the correct mutation by Sanger sequencing (Eurofins).

IGF2BP2 gene exon 10 insertion

By synthesizing *IGF2BP2* cDNA from total RNA isolated from HepG2 cells, we obtained *IGF2BP2* isoform B, which lacks exon 10 (129 nt). To generate the full-length *IGF2BP2* isoform A cDNA, we synthesized exon 10 *de novo* using overlapping oligonucleotides (Supplementary Table S1) and Q5® high-fidelity DNA polymerase (New England Biolabs). The *IGF2BP2* isoform B plasmid was PCR-linearized at the junction between exon 9 and exon 11 (Supplementary Table S1), and the synthesized exon 10 was then inserted into this site using Gibson assembly (New England Biolabs). The resulting plasmid of *IGF2BP2* isoform A was transformed into One Shot™ TOP10 Chemically Competent *E. coli* cells (Thermo Fisher Scientific), and colonies were screened by PCR and validated by Sanger sequencing (Eurofins).

RNA substrate preparation

Exactly 2.4 μg of total RNA was fragmented in 20 μl aliquots at 95°C for 4.5 minutes in a buffer containing 80 mM Tris–HCl (pH 8.5) and 8 mM $MgCl_2$. Fragmentation reactions were stopped by placing the aliquots on ice and adding 2 μl of 0.5 M EDTA. The aliquots were then pooled. Fragmented RNA was purified as follows: 0.5 volumes of RNA Clean® XP beads (Beckman Coulter), relative to the total volume of RNA, were pelleted on a magnetic stand and the supernatant was removed. The bead pellet was resuspended with the fragmented RNA solution and, immediately after, 1 volume of isopropanol was added to the bead–RNA mixture and incubated at room temperature for 10 minutes. Beads were pelleted, the supernatant was discarded, and beads were washed twice with 80% ethanol. The beads were air-dried at room temperature for 2 minutes and resuspended in 0.25 volumes of nuclease-free water. Next, RNA was 3′ dephosphorylated to a final concentration of 200 ng μl^{-1} in a reaction containing 100 mM Tris–acetate (pH 6), 10 mM $MgCl_2$, and 2.5 mM dithiothreitol (DTT), supplemented with 0.5 U μl^{-1} T4 PNK (New England Biolabs) and 2 U μl^{-1} RNase Inhibitor (Thermo Fisher Scientific). The reaction was incubated at room temperature for 4 hours. Dephosphorylated RNA was purified using the same RNA Clean® XP bead-based method (Beckman Coulter) as described before. 5′-end phosphorylation was performed using T4 PNK in 1× T4 PNK Buffer (New England Biolabs), supplemented with 1 mM ATP and a final RNA concentration of 100 ng μl^{-1}. The reaction was incubated at 37°C for 1 hour. RNA was again purified using the same bead-based procedure as described before, quantified, aliquoted (175 ng aliquots, 50 ng μl^{-1}), and stored at −80°C until further use. RNA quantification of fragmented RNA was performed using the Qubit™ microRNA Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For quality control, 2 ng of

the fragmented RNA substrate was used to prepare an input library. The fragmented RNA substrate was considered adequately 5′-end phosphorylated and 3′-end dephosphorylated if the resulting small RNA library, amplified with 12–15 PCR cycles, showed a uniform size distribution with a median insert size of approximately 40 nt, as measured by Agilent 2100 Bioanalyzer High Sensitivity DNA chips.

In vitro transcription

RBP-Halo plasmids were linearized using either SnaBI or FspI (New England Biolab) and subsequently gel-purified. A 1 μg aliquot of the linearized and purified plasmid was used for *in vitro* transcription using the MegaScript™ T7 Transcription kit (Thermo Fisher Scientific) at 37°C for 3 hours, following the manufacturer's instructions. After a 1 to 5 dilution, all *in vitro* transcription reactions were quantified by a NanoDrop 2000c spectrophotometer. Aliquots of 12 μg (1 μg μl^{-1}) were stored at −80°C until further use.

In vitro translation

A total of 10 μg of the *in vitro* transcribed RNA was *in vitro* translated using the Wheat Germ Extract *in vitro* Translation Kit (Promega) in a 50 μl final volume at 25°C for 2 hours, according to the manufacturer's instructions. The entire translation product was used for downstream assays, except for HUR titration experiments, in which 50% and 25% of the translation product were used.

Visualization of RBP-Halo fusions

Between 1 μl (for low molecular weight RBP-Halo fusions, <70–80 kDa) and 20 μl (for high molecular weight RBP-Halo fusion proteins, >80 kDa) of the *in vitro* translation reaction were incubated with HaloTag® Alexa Fluor® 488 Ligand (Promega) at a final dilution of 1:1000 at room temperature for 30 minutes. Afterwards, the reaction was mixed with an equal volume of Novex™ Tris–Glycine Sodium Dodecylsulfate Sample Buffer (Thermo Fisher Scientific) and denatured at 70°C for 10 minutes. Samples were loaded onto Novex™ WedgeWell™ 4–20% Tris–Glycine protein gels and electrophoresed at 225 V for 45 minutes. Fluorescent visualization was performed using a Bio-Rad ChemiDoc imaging system with settings for Alexa Fluor 488 detection.

RNA affinity purification

A total of 25 μl of Magne™ HaloTag® Ligand beads (Promega) were washed three times in PBSN [1× phosphate-buffered saline (PBS; Merck), 0.005% NP-40 (Thermo Fisher Scientific)], and then resuspended in 50 μl of PBSN. The resuspended beads in 50 μl of PBSN were then mixed with 50 μl of the *in vitro* translation reaction. To remove translation templates, 1 μl of RNase A (Thermo Fisher Scientific) and 1 μl of TURBO DNase™ (Thermo Fisher Scientific) were added. Bead-mediated capture of RBP-Halo fusions was performed at room temperature for 1 hour with end-over-end rotation. After binding, the beads were washed three times in 200 μl of PBSN supplemented with 1 M NaCl (high-salt washing buffer). The third high-salt washing step was performed by rotating end-over-end at room temperature for 10 minutes. Subsequently, beads were washed three more times with PBSN to remove excess salt by rotating end-over-end at room temperature for 10 minutes. The RBP-Halo-bound beads were resuspended in 50 μl of PBSN and combined with 150 ng of the fragmented RNA substrate in 50 μl. While washing the beads, a 3 μl aliquot of fragmented RNA substrate (50 ng μl^{-1}) was thawed on ice, denatured at 70°C for 2 minutes, immediately placed on ice, and diluted with 47 μl of ice-cold PBSN. The denatured and diluted RNA was then refolded at room temperature for 45 minutes before mixing with RBP-Halo beads. The binding reaction was carried out at room temperature for 1 hour by rotating end-over-end. Afterwards, beads were washed three times with 200 μl of PBSN by pipetting up and down vigorously 20 times. Beads were then resuspended in 15 μl of nuclease-free water. Elution was done by incubating at 70°C for 8 minutes, pipetting the reaction up and down three times while the PCR tube remained in the thermocycler. Immediately afterwards, the reaction tube was placed on a magnetic rack and 11 μl of the supernatant was transferred to a new tube on ice for library preparation. A minimum of two replicates for each protein are recommended.

RAP-qPCR

A total of 11 μl of eluate from an RNA affinity purification reaction was mixed with 1 μl of random hexamers (250 ng μl^{-1}) (Thermo Fisher Scientific) and 1 μl of 10 mM dNTPs (Thermo Fisher Scientific). The mixture was incubated at 65°C for 5 minutes and then placed on ice for 5 minutes. Separately, 1 μl of SSII was combined with 1 μl of 100 mM DTT, 1 μl of RNase inhibitor, and 4 μl of 5× First Strand buffer (Thermo Fisher Scientific), and added to the denatured and annealed RNA–hexamer mix. Reverse transcription was performed with an initial incubation at 25°C for 20 minutes, followed by 45 minutes at 42°C and heat inactivation at 90°C for 10 minutes. For quantitative PCR (qPCR), 1 μl of the reverse transcription reaction was mixed with 10 μl of PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific), 1 μl of 10 μM gene-specific primers (Supplementary Table S1), and nuclease-free water to a final reaction volume of 20 μl. When designing primers for RAP-qPCR, an optimal amplicon size is between 70 and 80 nt. RAP-qPCR was performed using a QuantStudio™ 5 Real-Time PCR System with the following cycling conditions: 50 cycles of 95°C for 15 seconds, 50°C for 30 seconds, and 72°C for 30 seconds. Relative standard curves and C_t values were obtained using Applied Biosystems™ Analysis Software (Thermo Fisher Scientific). Fold changes relative to HaloTag control were calculated using the $\Delta\Delta C_t$ method.

RNAseq library preparation

RAPseq libraries were prepared using the NEXTflex® Small RNA-Seq Kit v3 (Bioo Scientific) with two modifications to the manufacturer's protocol. First, MMLV (Moloney murine leukemia virus) reverse transcriptase was replaced with SSII (Thermo Fisher Scientific). Second, the reverse size selection step was omitted and preparation proceeded by cleaning up the reverse transcription reaction. After 12 PCR amplification cycles, 1 μl of the PCR product was analyzed using Agilent 2100 Bioanalyzer High Sensitivity DNA chips according to the manufacturer's instructions. Libraries that were insufficiently amplified after 12 cycles underwent an additional two PCR amplification cycles. Adapters and primers were removed using AMPure® XP beads (Beckman Coulter) with a bead-to-sample ratio of 1.5×, according to the manufacturer's instructions. Final library size distribution was assessed with Agilent

2100 Bioanalyzer High Sensitivity DNA chips, and concentrations were measured using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific).

mod-RAPseq

A total of 2 ng of the RAPseq input library was PCR-amplified using primers F_T7_RAPseq and R_T7_RAPseq (Supplementary Table S1) and Q5® DNA Polymerase (New England Biolabs) for 20 cycles. The PCR product was purified using AMPure® XP beads (Beckman Coulter) at a 2.0× bead-to-sample ratio, following the manufacturer's protocol. Next, 500 ng of the purified PCR product was used as the template for *in vitro* transcription using the MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific). Transcription was performed at 37°C for 4 hours. The *in vitro* transcription reaction was then treated with 1 µl of TURBO DNase™ (Thermo Fisher Scientific) and incubated at 37°C for 1 hour to remove DNA templates. RNA was purified using RNA Clean® XP beads (Beckman Coulter) at a 1.5× ratio as per manufacturer's instructions. The final RNA product was quantified using the Qubit™ microRNA assay kit (Thermo Fisher Scientific), diluted to 50 ng μl^{-1}, and stored in 3 µl aliquots at −80°C. RNA affinity purification was performed as described. For reverse transcription, 11 µl of RAPseq eluate was mixed with 1 µl of 10 µM R_T7_RAPseq primers in a 20 µl reaction containing SSII (Thermo Fisher Scientific), and incubated at 42°C for 45 minutes. The reaction was heat-inactivated at 90°C for 10 minutes. cDNA was purified according to the NEXTflex® Small RNA-Seq Kit v3 (Bioo Scientific) cleanup procedure. Purified cDNA was then PCR-amplified in a 25 µl reaction using 0.5 µl of Q5® DNA polymerase (New England Biolabs), 1 µl of universal primer, and 1 µl of barcoded primer (both from the NEXTflex® Small RNA-Seq Kit v3, Bioo Scientific). After six cycles of amplification, the PCR product was purified using AMPure® XP beads (Beckman Coulter) using a 1.5× bead-to-sample ratio.

co-RAPseq

A total of 25 µl of Magne® HaloTag® Ligand beads was washed three times with PBSN. After removing the storing solution, the beads were resuspended with 100 µl of a mixture containing two combined 50 µl *in vitro* translation reactions for HUR and PTBP1, respectively. RNA affinity purification and library preparation were performed as previously described.

High-throughput sequencing

RNA libraries were sequenced either on an Illumina® NovaSeq using 300-cycle paired-end reads or on a NextSeq500 and NextSeq2000 using 83-cycle paired-end reads. Considering the RAPseq substrate size, generating 80- to 100-cycle paired-end reads is sufficient for downstream analyses.

Read processing and alignment

FASTQ files were first trimmed to remove NEXTflex adapters from both mates by cutadapt (v3.1). Unique molecular identifiers (UMIs), located in the first four bases of read 1 and read 2, were extracted and appended to the read names. Reads aligning to a collection of rRNA and tRNA genes were removed by using HISAT2 (v2.2.1) [25]. Remaining paired-end reads were aligned to either the GRCh38 human genome or the GRCz11/danRer11 zebrafish genome using HISAT2. Only uniquely mapping reads were retained. PCR duplicates were removed using UMI-based deduplication with umi_tools (v1.0.0) [26]. Read processing statistics are provided in Supplementary Table S3.

Binding site identification

Binding sites were identified using MACS2 (v2.2.6) subcommands [27]. Briefly, read pileups were computed with bedtools (v2.29.2) [28], normalized to the same sequencing depth as the HaloTag control library. The HaloTag control signal was then subtracted from both the RBP and input control libraries using macs2 bdgcmp. A second subtraction removed the remaining signal present in the input control. Candidate genomic regions with enriched signal, relative to both HaloTag and input control libraries, were identified from the remaining RBP read pileups using macs2 bdgpeakcall. A candidate region was retained for further processing if it had at least five read pair counts and no significant enrichment in the HaloTag over input control (fold change (FC) < 2 and a false discovery rate (FDR)-adjusted $P < 0.05$) in both replicates. Candidate genomic regions passing all filters were considered significant binding sites. For each binding site, a peak binding score (BS) was computed, which represents the mean peak fold change weighted by the base 2 logarithm of the peak normalized read counts:

$$BS_{peak} = \frac{1}{4}\left[\begin{array}{l}\left(log_2\left(Rep1\right) * \left(\frac{Rep1}{FC_H} + \frac{Rep1}{FC_I}\right)\right) \\ +log_2\left(Rep2\right) * \left(\frac{Rep2}{FC_H} + \frac{Rep2}{FC_I}\right)\end{array}\right]$$

Where FC_H and FC_I represent fold changes over Halo and over input, respectively.

Comparative binding site analysis

To compare HUR orthologs and IGF2BPs variants, binding site metafiles were built by overlapping the peak coordinates of orthologs or variants, respectively, using bedtools merge (v2.21.0). For each RBP, the signal within the overlapping regions was retrieved from each replicate of the respective ortholog or variant.

Binding site annotation

Binding sites were annotated using Gencode v37 GTF (for human substrate assays) and Ensembl GRCz11 GTF (for zebrafish assays) files and the R package GenomicFeatures [29]. Only binding site summits (1-nucleotide positions) were used for annotation to avoid multiple feature assignments. The findOverlaps function of GenomicFeatures (v1.54.4) was used to intersect the binding site summit coordinate with the GTF file, and only summits that overlapped were considered. When multiple features were assigned to one feature, annotation priority was assigned based on genomic context as follows: intron < non-coding RNA (ncRNA) < 5′-untranslated region (UTR) < 3′UTR < coding sequence (CDS).

De novo motif identification

The DREME algorithm of the MEME Suite package (v5.1.1) [30] was used to identify enriched motifs within 50 nt bins for HURs or 60 nt bins for HNRNPA1, HNRNPC, PTBP1, RBFOX2, and YBX3, centered around the binding site summits, extracted from the annotated peak file. A background

set of sequences representing unbound control sites was taken 1000 nt downstream (3′ direction) from each peak, matching the bin sizes of the positive sequences. Background sequences were provided to DREME as a custom background model or, in some cases, the default DREME shuffled dinucleotide background model was used. For each RBP, optimized DREME arguments were used for motif identification.

Differential binding analysis

The edgeR (v4.0.16) functions glmQLFit and glmQLFTest [31] were used to identify differential binding events with an FC > 2 and FDR-adjusted $P < 0.05$. HUR- and PTBP1-binding sites located within 30 nt of each other were considered. edgeR was used to compare normalized read counts across both HUR and PTBP1 replicates. First, binding differences were computed by summing all normalized binding site counts for each individual gene; these per-gene summed counts were then used as input for edgeR. For IGF2BP variants, all binding sites per paralog and the corresponding variants were first merged, keeping the fold change for each paralog. The resulting paralog metafile was used to identify differentially bound sites.

Gene Ontology and reactome pathway enrichments

clusterProfiler (v4.10.1 and v4.12) was used to identify significantly enriched Gene Ontology (GO) terms using enrichGO, or Reactome Pathways using enrichPathway [32] with org.Hs.eg.db (v3.18.0 and v3.19.1). Enrichments were determined at an FDR-adjusted $P < 0.05$ or $P < 0.01$, over the default package backgrounds or the expressed HepG2 transcriptome according to our input library (Supplementary Table S4). GO terms through which HUR and PTBP1 exerted cooperative functions were tested by considering genes that were differentially bound or only bound in the co-RAPseq assay. The background gene set consisted of the union of all genes bound in all three assays.

Benchmarking of motif identification among different methods

To compare the motifs identified by our method with those obtained from other approaches (eCLIP-seq, RNAcompete, and RBNS), we profiled RBP–RNA interactomes using HepG2 RNA. eCLIP-seq data were retrieved from the ENCODE Project (https://www.encodeproject.org/) with the following accession numbers: HNRNPA1 (ENCFF797GSK), HNRNPC (ENCFF440ROZ), PTBP1 (ENCFF726SQU), RBFOX2 (ENCFF871NYM), and YBX3 (ENCFF185OEI). RBNS data were also obtained from the ENCODE Project, with accession numbers HNRNPC (ENCSR569UIU) and RBFOX2 (ENCSR441HLP). RNAcompete data were retrieved from Ray *et al.* [22].

We recomputed the position weight matrices for both RAPseq and eCLIP-seq using the DREME tool (see "De novo motif identification"). Read counts were obtained from BAM files and normalized to sequencing depth for each replicate and size-matched input control. Bed files were annotated using the same pipeline applied to RAPseq peaks (see "Binding site annotation''), computing fold changes and read densities, and performing *de novo* motif identification. Because position weight matrices for RBNS were not available, motif files were downloaded from the ENCODE Project. RNAcompete position weight matrices were obtained from the corresponding publication.

Integrated analysis of RBP–RNA interactomes

A set of 37 RBPs was selected for RAPseq analysis and subsequent integration after intersecting data from four independent studies [2–4, 33]. Established cRBPs and novel ncRBPs were selected according to their domain composition and canonical functions. Only RBPs with >1000 binding sites in the transcriptome were selected for downstream analysis. For the integration of RNA targets, overlapping and unique target sets for the different RBPs were limited to a minimum of 20 genes and visualized as a network using Cytoscape (v3.9.1) [34]. To investigate the effects of RNA-binding on RNA levels, we examined publicly available RNAseq datasets of small interfering RNA (siRNA)-mediated knockdown experiments of YBX3, PTBP1, RBFOX2, and STMN1 [35, 36]. We assessed the enrichment of RBP–target sets in the knockdown experiment results with gene set enrichment analysis (GSEA) using the fgsea R package (https://github.com/alserglab/fgsea). Gene lists were ranked according to signed $\log_{10}P$-values, and the resulting normalized enrichment score values were extracted for assessment. GO over-representation analysis with clusterProfiler was performed for selected gene target sets of YBX3, PTBP1, RBFOX2, and STMN1 to identify relevant biological processes, applying a lenient cut-off of $P < 0.1$. The rrvgo R package was used to collapse over-represented terms according to their similarity using the "Rel" method and a threshold of 0.9 to define GO clusters [37]. Normalized gene expression values for patients with liver cancer in the TCGA cohort were retrieved from the LinkedOmics platform [38]. Survival analysis results of patients with liver cancer separated by high and low RNA levels of selected genes were retrieved from the GEPIA2 database [39].

Results

RAPseq captures structure-, sequence-, and modification-specific RBP–RNA interactions using native transcriptomes

RAPseq is a binding assay that measures interactions between a recombinant protein and native RNA using high-throughput sequencing as a readout (Fig. 1A). To generate the RNA substrate, we chemically fragmented total RNA isolated from cells and tissues to a product with a median size of 35 nt [interquartile range (IQR) 25–49 nt] (Supplementary Fig. S1A). We subsequently performed 3′-end dephosphorylation and 5′-end phosphorylation to enable ligation of sequencing-compatible adapters (Fig. 1A). This RNA preparation resulted in substrate uniformity and allowed for efficient amplification and sequencing. For the production of the recombinant RBP, we cloned the respective cDNA fused to HaloTag [40] into a cell-free expression vector (Fig. 1A). C-terminal Halo tagging ensured that we assayed full-length rather than N-terminally tagged truncated fusions (Supplementary Fig. S1B, C). We treated the RBP from the *in vitro* expression system with RNase to eliminate residual RNA (Supplementary Fig. S1D) and removed contaminants by high-salt washes. After purification, the RBP-Halo fusion was incubated with the fragmented RNA substrate, and RBP-bound RNAs were recovered and deep sequenced (Fig. 1A) (see Materials and methods). We compared the binding efficiency of fragmented versus

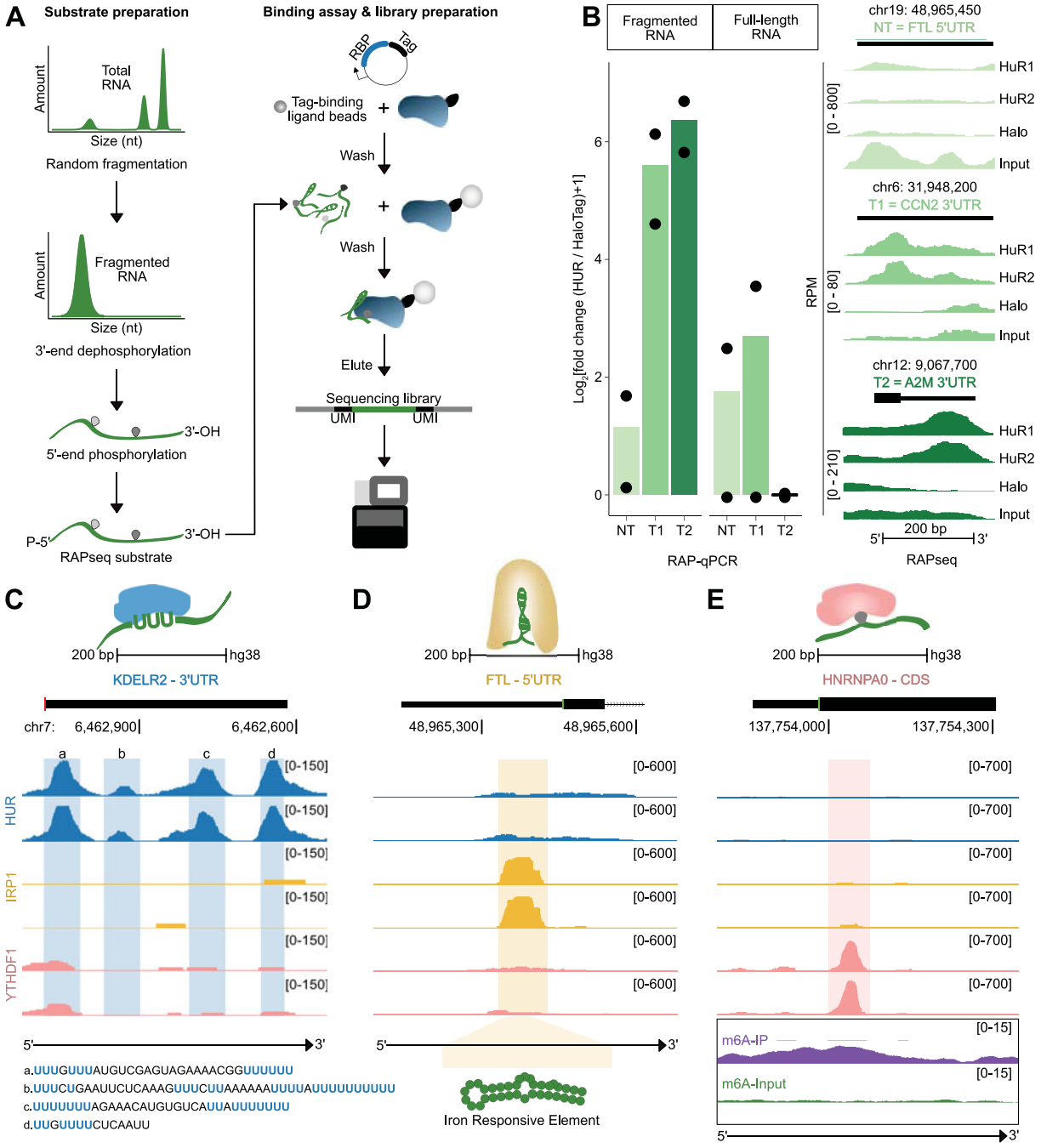


Figure 1. RAPseq captures RBP-binding facilitated by RNA structure, sequence, and modification. (**A**) Schematic presentation of the experimental setup of RAPseq. RAPseq substrates are generated from native total RNA (left). Production of an RBP-Halo fusion allows protein purification and RNA-binding assay. Next-generation sequencing serves as readout using UMIs (right). (**B**) The dot plot displays binding of HUR to non-targeting (NT) and targeting (T1 and T2) RNA regions (*x*-axis) of fragmented (left) and full-length (right) RNA from HepG2 cells. RAP-qPCR (left) was performed in two replicates (colored dots), and means (black horizontal lines) are shown. RNA-binding is displayed as $\log_2$-scaled fold change ($\Delta\Delta$Ct) of HUR over the -HaloTag control (*y*-axis). RAPseq coverage tracks (right) represent HUR-binding to fragmented RNA [in reads per million (RPM)]. Genomic locations of the corresponding target regions (T1 and T2) are specified. Tracks for two HUR replicates and two RAPseq controls (HaloTag and RNA input) are shown. (**C–E**) Genome tracks demonstrate binding of (C) HUR to ARE and GRE motifs, (D) IRP1 to the iron-responsive element, and (E) YTHDF1 to a modified nucleotide (m^6A) within *KDELR2*, *FTL*, and *HNRNPA0* mRNAs, respectively, in HepG2 cells. Genomic locations with a scale bar indicate the length of the genomic region in bases, and gene features (black rectangle, exon and UTR; gray line, intron; arrow, direction of transcription) (*x*-axis) and normalized read density (RPM, *y*-axis) are shown. Vertical lines highlight bound RNA elements. Tracks for two RBP replicates and m^6A-specific RNA immunoprecipitation (IP, purple) over input control (green) are presented.

full-length total RNA substrates and observed that fragmentation improved the RBP–RNA pulldown efficiency (Fig. 1B), probably due to improved accessibility of RBP-binding sites.

To assess the dynamic range and sensitivity of RAPseq, we performed titration experiments using three concentrations of the RBP HUR, i.e. the total *in vitro* expression reaction (100%) and diluted inputs at 50% and 25%. At 100% and 50% input, RAPseq identified 4129 and 3801 RNA targets, respectively, of which 2790 were shared. The 25% titration yielded 1543 targets, with up to 86% (1324/1543) of targets overlapping with the higher input titrations. In total, 411 RNA targets were common to all the three HUR RAPseq titrations and the corresponding eCLIP dataset (Supplementary Fig. S1E). This result suggests a threshold effect in detection sensitivity that correlates with RBP input concentrations. We evaluated the quantitative performance of RAPseq by computing a peak binding score, which we defined as the average enrichment of reads at a given peak over our two negative controls, weighted by the average replicate-normalized read counts (see Materials and methods). Higher peak binding score values were consistently observed among targets shared across the three titrations, while unique targets exhibited significantly lower scores (Supplementary Fig. S1F). We therefore chose the highest concentration of RBP input for all subsequent assays.

To assess whether RAPseq captured known RBP-binding modalities to RNA, we profiled three RBPs that have sequence- (HUR) [41–43], structure- (IRP1) [44], and modification-specific (YTHDF1) [45, 46] substrate requirements. In line with previous studies, we detected a cluster of AU- and GU-rich RNA elements (ARE, GRE) in the 3′UTR of the *KDELR2* mRNA as HUR-binding sites (Fig. 1C). IRP1 specifically bound the iron-responsive element in the 5′UTR of the *FTL* mRNA (Fig. 1D). For YTHDF1, we identified binding sites overlapping with known m^6A-modified sequences in the *HNRNPA0* mRNA [6] (Fig. 1E).

In sum, RAPseq is the first *in vitro*-based method capable of capturing a wide range of interactions between RBPs and native cellular RNA, involving sequence, structure, and RNA modifications.

RAPseq recapitulates known RBP-binding specificities with high quantitative accuracy

To benchmark RAPseq, we focused on five cRBPs (HNRNPA1, HNRNPC, PTBP1, RBFOX2, and YBX3) that recognize RNA through sequence-specific interactions. We profiled their RNA interactomes using HepG2 RNA and compared the results with publicly available *in cellulo* eCLIPseq data [36] and the *in vitro* RBP–RNA interaction assays RNAcompete [22] and RBNS [23]. For the five RBPs, we observed high similarity in the bound motifs across the diverse interaction assays (Fig. 2A; Supplementary Fig. S2A). For the well-characterized RBP RBFOX2, we observed a subtle deviation from the canonical GCAUG motif. We found comparable frequencies of cytosine (C) and adenine (A) at the second position in the motif, forming the GAAUG motif, consistent with previous reports [47]. Using lenient motif criteria (see Materials and methods), we identified a more generalizable RBFOX2 motif, GNWYG (where N corresponds to any nucleotide, W to A/U, and Y to C/U) in nearly 90% of binding sites, compared with 44% for the more specific G(C/A)AUG. Notably, binding sites containing only the canonical GCAUG motif had significantly higher peak binding score values than the other 13 possible GNWYG 5-mers, confirming GCAUG as the motif with the highest binding affinity (Supplementary Fig. S2B). All GNWYG 5-mers were specific to RBFOX2 compared with the other RBPs assayed by RAPseq (Supplementary Fig. S2C). Peak binding scores for the majority (86%, 12/14) of the GNWYG motif variants increased upon inclusion of an additional 5′U (Supplementary Fig. S2D), validating prior findings on enhanced binding [23]. Furthermore, binding sites containing two or more 5-mers exhibited even stronger binding than sites with a single motif (Supplementary Fig. S2D). We also confirmed the generalized GNWYG motif in eCLIP data when examining enrichment based on motif frequencies, excluding GCAUG instances (Supplementary Fig. S2E). Similarly, we observed increased YBX3-binding scores in the presence of additional 5′As flanking the previously described core CAHC motif (where H corresponds to a U, C, and A) [48] (Supplementary Fig. S2F).

We calculated peak binding score values to evaluate the quantitative capacity of RAPseq (see Materials and methods). The peak binding score values in the HNRNPC assays were positively correlated with the uracil content of RNA targets [49, 50]. The binding strength plateaued when the uracil content exceeded 50%, forming a characteristic biochemical binding curve that was not detectable in eCLIP data (Fig. 2B). We next compared the concordance between RAPseq and eCLIPseq regarding motif location and read enrichment of the five RBPs. In RAPseq, the motif consistently mapped to the center of peak regions, resembling profiles from transcription factor chromatin immunoprecipitation (ChIP) assays. This allowed the use of standard peak callers (see Materials and methods). In contrast, iCLIP and eCLIP often display a positional shift between motif and read coverage, requiring assay-dependent customized algorithms [51] (Fig. 2C, D). RAPseq showed higher reproducibility and significance of binding compared with eCLIPseq (Spearman rank correlation for RAPseq $\rho = 0.93$–0.94 and for eCLIPseq $\rho = 0.58$–0.92, two-tailed Wilcoxon rank sum test for eCLIPseq and RAPseq, $P < 0.01$) (Fig. 2E, F). When combining peaks on the gene level, RAPseq identified a larger set of RBP-bound genes. However, the majority (56–83%) of RBP-bound genes in eCLIPseq overlapped with RAPseq, suggesting that many *in cellulo* interactions are governed by intrinsic binding specificities (Fig. 2G). Furthermore, transcripts bound *in vitro* were significantly stronger bound *in cellulo* (two-tailed Wilcoxon rank sum test, $P < 0.01$) (Fig. 2G), confirming the biological relevance of RAPseq-detected interactions.

Overall, RAPseq quantitatively captures RBP-binding specificities and enables reliable profiling of RBP–RNA interactomes.

Systematic profiling of RBP targets reveals specialized RNA interactomes

Despite the increased identification of ncRBPs through recent proteomic- and sequencing-based technologies, their RNA-binding mechanisms remain limited [52]. To evaluate whether RAPseq can discover RNA targets of ncRBPs, we intersected RBPomes from four independent studies [2–4, 33]. Across these datasets, one-third (243/716) of the common RBPs were ncRBPs (Supplementary Fig. S3A). We selected 26 ncRBPs representing diverse domain composition and canonical functions based on pre-defined criteria including protein size suit-

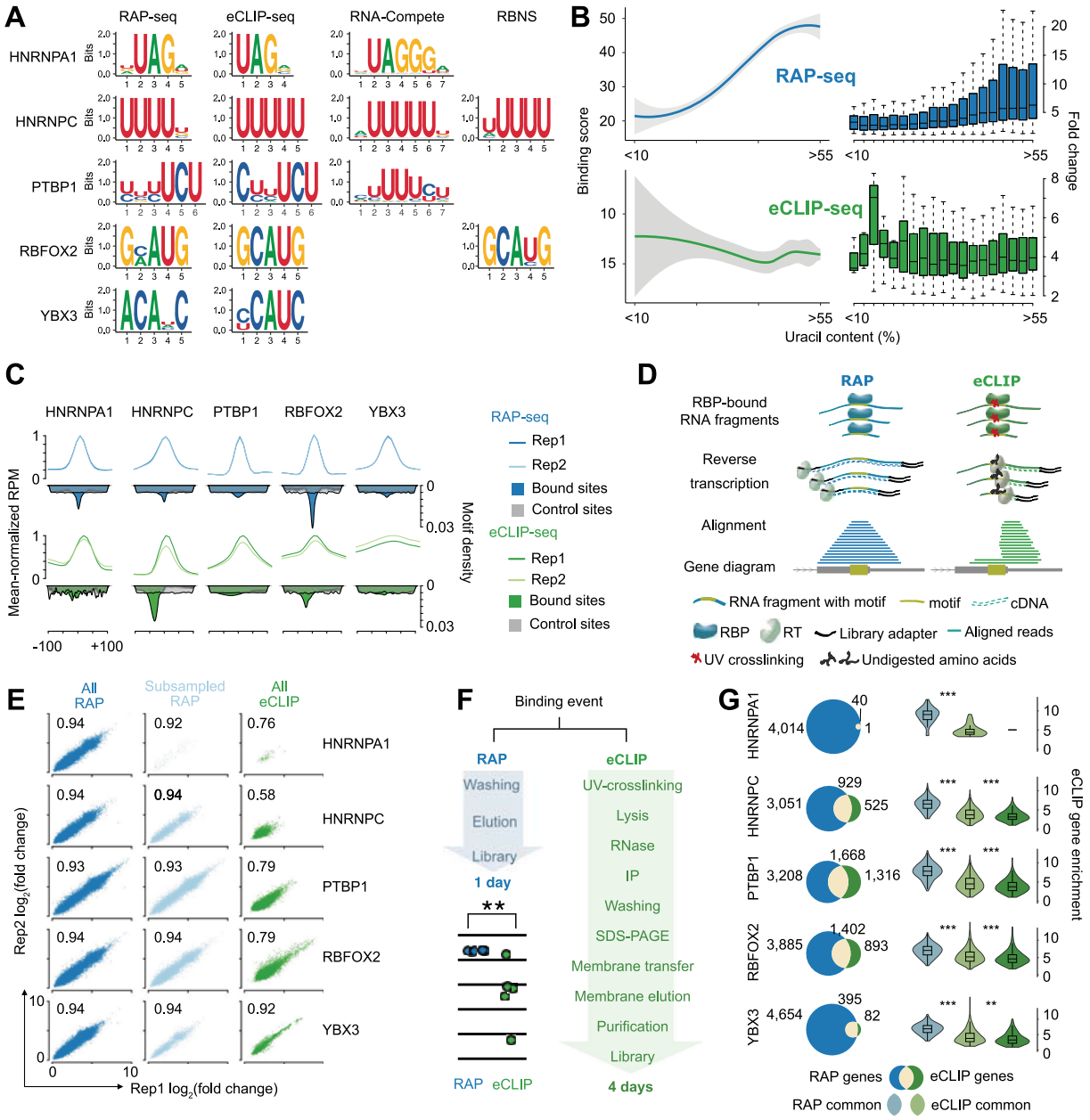


Figure 2. Accuracy and robustness of RAPseq are confirmed by benchmarking against state-of-the-art methods. (**A**) Position weight matrices display *de novo* identified motifs of five RBPs with two replicates each in RAPseq and eCLIPseq assays and published motif models from *in vitro* assays (RNAcompete and RBNS). (**B**) LOESS-fitted smoothed regression lines (left) and box plots (right) of HNRNPC RAPseq (blue, top) and eCLIPseq (green, bottom) assays show HNRNPC-binding in terms of peak binding scores (left) and fold changes (right) (*y*-axes) as a function of uracil content (*x*-axes) in a 40 nt bin centered around the peak summit (for RAPseq, top) and around the highest motif density location (for eCLIPseq, bottom), $n = 2$. (**C**) Line and density plots compare the motif locations (polygons) with respect to the average normalized read enrichment (RPM, lines) for RAPseq and eCLIPseq for five RBPs. Line plots show the average RPM for all peaks plotted as a fraction of the maximum mean RPM for two experimental replicates (top, *y*-axes). The bin size for the line plots is 1 nt. Density plots display the density of RBP-binding motif locations (A) at bound sites for RAPseq (blue), eCLIPseq (green), and the same motifs over unbound control sites (gray). The *x*-axes show the distance from the peak summit. Ticks are drawn at −100, −50, 0, 50, and 100 nt from the respective RBP peak summit. (**D**) Illustrative model explains the read enrichment at (RAPseq, left) and next to the motif (eCLIPseq, right) due to intrinsic experimental differences when capturing RBP–RNA interactions. (**E**) Scatter plots show the correlations in binding enrichments ($\log_2$ of peak fold changes) between two replicates (*x*,*y*-axes) for all RAPseq binding events (left column) and subsampled RAPseq binding events (middle column) to match the number of eCLIP binding events (right column) for five RBPs. Spearman's rank correlation coefficients are indicated. (**F**) Comparison of the experimental steps in RAP (left) and eCLIP (right) protocols. The dot plot correlates the mean of two replicates for five RAPseq (blue) and eCLIPseq (green) experiments, derived from (E). Asterisks indicate statistical significance (Spearman rank correlation, $**P < 0.01$). (**G**) Two-way Venn diagrams (left) intersect the number of genes with identified binding sites for five RBPs in RAPseq (blue) and eCLIPseq (green), two replicates each. The intersected area is colored (yellow) and the number of genes per intersection is displayed. Violin plots (right) display gene enrichments ($\log_2$ scale) for RAPseq common genes (light blue), eCLIPseq common genes (light green), and eCLIPseq only genes (green). Asterisks represent statistical significance (two-tailed Wilcoxon rank sum test, $**P < 0.01$, $***P < 0.001$). The gene enrichments are computed as the sum of all peak enrichments (binding score) per gene.

able for cloning, domain composition, binding specificity, and functional annotations. Their RNA interactomes were profiled using RAPseq, alongside 11 cRBPs as controls. All profiled RBPs are listed in Supplementary Table S5, which includes protein lengths, domain compositions, intrinsically disordered regions, and predicted RNA-binding motifs. While cRBPs bound to a median of 7297 binding sites across the transcriptome, ncRBPs exhibited fewer interactions, with a median of 560 binding sites (Supplementary Fig. S3B). However, 11 ncRBPs displayed substantial binding activity, each interacting with >1000 RNA sites (Supplementary Fig. S3C). Among them, HSPA8 and PEBP1 showed the largest number of binding events (4810 and 4735, respectively), whereas CCDC124 and CCT2 had the fewest (27 and 26, respectively) (Supplementary Fig. S3C). On average, cRBPs had over two (range: 1.03–2.93) binding sites throughout the gene body, whereas ncRBPs exhibited approximately one binding site per target gene (Supplementary Fig. S3D). Moreover, ncRBPs demonstrated weaker binding strengths compared with cRBPs, even after correcting by the total number of peaks (Supplementary Fig. S3E). This trend was maintained when investigating high binding strength ncRBPs with over 1000 peaks (Supplementary Fig. S3F). Notably, STMN1 showed the highest binding strength among ncRBPs despite having relatively few binding sites, suggesting selective, high-affinity interactions with functionally relevant RNAs and a specialized role in post-transcriptional regulation (Supplementary Fig. S3C, F).

To assess the specialization for each of the generated RBP–RNA interactomes, we focused on the high binding strength RBP group, consisting of 10 cRBPs and 11 ncRBPs, and analyzed their bound target gene sets (Fig. 3A). The number of binding sites correlated with the number of target genes, but not the binding strength of RBPs (Fig. 3A; Supplementary Fig. S3C, F). Notably, the binding strength observed for ncRBPs was generally low compared with cRBPs, with only some exceptions displaying comparable binding capability. Several RBPs displayed distinct binding preferences across gene regions. For instance, PTBP1 showed higher binding at introns and UTRs, but had a lower fraction of binding sites in coding sequences compared with other RBPs. HUR displayed a preference for binding 3′UTRs whereas STMN1 showed the highest differences in 5'UTRs among ncRBPs.

We further evaluated the specialization of RBPs by building a unified network of RNA interactomes (Fig. 3B). This approach revealed unique sets of target genes for the different RBPs as well as co-targeted transcripts. Specifically, YBX3, PTBP1, RBFOX2, and STMN1 formed a dense subnetwork of co-regulated targets (Fig. 3B). To assess the regulatory consequences of binding, we examined RNA abundance changes using publicly available RNAseq data from RBP knockdown experiments [35, 36]. Target genes of each RBP were stratified into sets according to the location of RBP-binding sites within the gene body and examined with GSEA for changes in gene expression. We found that PTBP1 and RBFOX2 showed the greatest changes in RNA levels compared with random gene sets when binding introns, while YBX3 and STMN1 affected transcripts through binding to 3′UTRs and exons of non-coding genes, respectively (Supplementary Fig. S3G). Although overall changes in RNA abundance were modest (Supplementary Fig. S3H), a subset of co-regulated transcripts, such as those shared by YBX3 and RBFOX2, were affected upon RBFOX2 knockdown (Fig. 3C). This suggests that only a fraction of RBP targets exhibit changes in RNA abundance, indicating alternative modes of post-transcriptional regulation independent of RNA level changes.

Lastly, after profiling the RNA interactomes in liver cancer cells (HepG2) and identifying of several RBPs in cancer development and progression, we examined the abundance of RBP–targets across hepatocellular carcinoma (LIHC) stages and their association with patient survival. Out of the 61 genes in the RBP subnetwork, 23 were significantly linked to LIHC progression and survival outcomes (Fig. 3D, columns 1–3). Functional enrichment analysis revealed key processes, with the most enriched processes involving genes targeted by multiple RBPs, such as cell–cell adhesion, regulation of lipid catabolism, and post-embryonic development (Fig. 3D, column 4). Additionally, some processes showed specificity for individual RBPs like AAAS, CEP131, NCKAP5L, and MTCL1, which were targeted by STMN1 and associated with microtubule formation and poor patient survival. Variations in the location of RBP-binding sites were observed among genes, with most sites mapping to coding sequences (Fig. 3D, column 5).

Together, our systematic analysis confirms the ability of RAPseq to profile diverse RBP–RNA interactomes. Our findings demonstrate that both cRBPs and ncRBPs form specialized and functionally distinct RNA interactomes, displaying unique binding preferences and regulatory capacities. The results suggest that RBPs, especially ncRBPs, may serve as crucial post-transcriptional modulators with potential roles in disease-relevant pathways, particularly impacting liver cancer progression and patient outcomes.

HUR-binding to uracil triplets is conserved but exhibits species-specific transcript processing preferences

Speciation leads to rapid changes in transcriptome regulation, requiring adaptation of RBP–RNA interactions. RAPseq is the first method enabling cross-species comparisons between RBP orthologs and native RNA substrates. RBPs encoded by the ELAV gene family are evolutionarily conserved, featuring three RNA recognition motifs (RRMs) with strong binding affinity for AU-rich elements [53]. Given that HUR is the most divergent paralog of the ELAV gene family that evolved through gene duplication in vertebrates [53], we quantified and compared HUR-binding specificities across vertebrate evolution. We inspected HUR orthologs from human, mouse, opossum, chicken, frog, and zebrafish. All orthologs had a similar length, and the amino acid sequence identity decreased gradually from human to zebrafish (Fig. 4A). Except for frog HUR, amino acid changes resided outside the RNA-interacting β-sheets of the three RRMs, suggesting minimal impact on RNA-binding (Supplementary Fig. S4A). Despite two amino acid differences in the RRM1 and RRM3 β-sheets of the frog HUR, existing crystallographic models indicated that uracil triplet (UUU)-binding is mediated by all three RRMs in human HUR, implying conserved interaction modes [43, 44]. To validate the conservation of binding specificity, we produced HUR proteins from the six orthologs and conducted RAPseq assays with the same human RNA substrate (Fig. 4B). We found that all orthologs bound to UUUs, which were significantly enriched in over 85% of all ortholog bind-

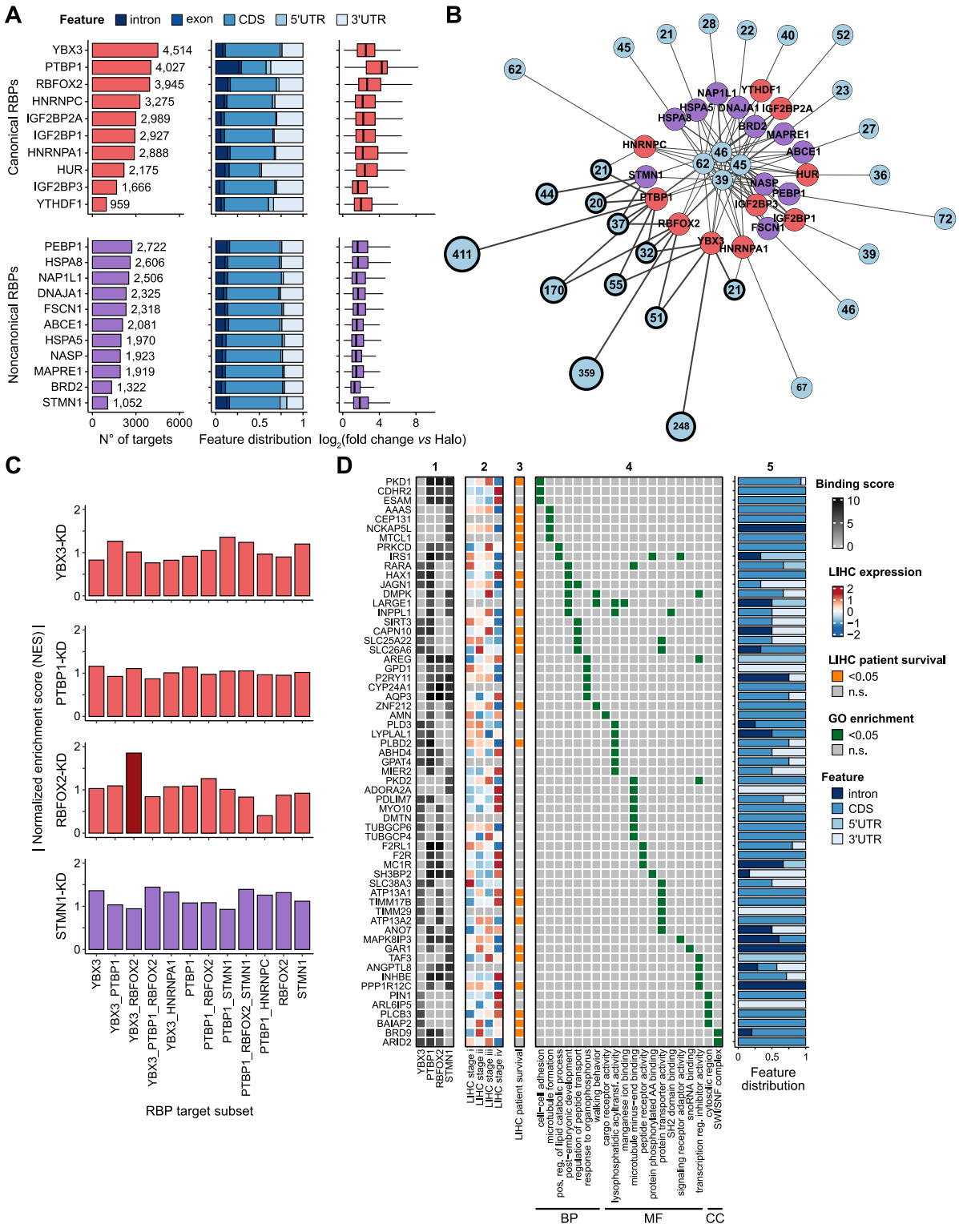


Figure 3. Combined analysis of RBP–RNA interactomes reveals binding and regulatory preferences. (**A**) Plots summarize the characteristics of cRBP (red) and ncRBP (purple), with two replicates each. Horizontal bar plots (left) display the number of gene targets per RBP. Stacked bar plots (middle) display the fraction of binding sites in different regions of the gene body. Box plots (right) illustrate the binding strength across the detected binding sites. Each box indicates the IQR, median (horizontal line), and 1.5× IQR (whiskers). (**B**) The RBP–RNA interactome network represents overlaps of RNA targets between cRBP (red nodes) and ncRBP (purple nodes), with two replicates each. The number of target genes in each of the overlapping and non-overlapping sets is indicated (blue node size). A subnetwork of targets from YBX3, PTBP1, RBFOX2, and STMN1 is highlighted (bold edges). The represented target sets are limited to a minimum size of 20 genes. (**C**) Bar plots show the NESs obtained from GSEA of RBP–target sets from the subnetwork highlighted in (B) using available RNAseq data from cRBP (red) and ncRBP (purple) knockdown experiments. The RNA–target subset co-regulated by YBX3 and RBFOX2 is highlighted (dark red). (**D**) Functional implications of gene targets from the subnetwork highlighted in (B) that are associated with an enriched GO term. For each target gene, the RBP-binding score (column 1), mRNA abundance in liver cancer (LIHC) stages (column 2), association with LIHC patient survival (column 3), association with enriched GO terms (column 4), and gene region distribution of binding sites (column 5) are displayed.

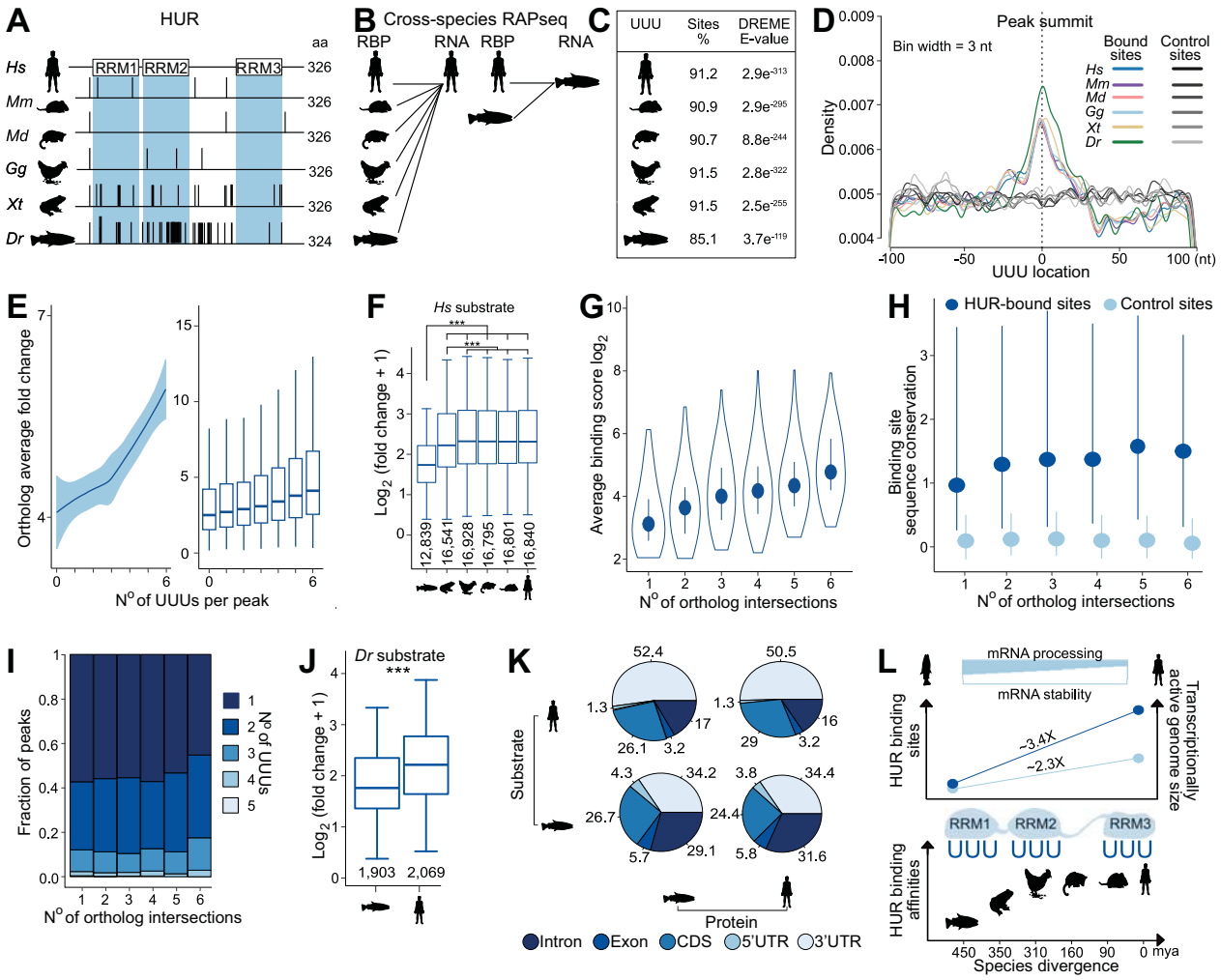


Figure 4. RNA-binding specificity of HUR is conserved in vertebrates but acquired functional differences. (**A**) Orthologous vertebrate HUR protein length and RRMs 1–3 (white boxes and vertical light blue lines) are shown. Amino acid changes (black lines) for each ortholog compared with the human reference are highlighted. *Hs*, human; *Mm*, mouse; *Md*, opossum; *Gg*, chicken; *Xt*, frog; and *Dr*, zebrafish. (**B**) Schematic representation of the performed RAPseq assays using vertebrate HUR proteins and either the human (left) or zebrafish RNA substrate (right). (**C**) The table lists the percentage and corresponding DREME E-value of UUU motifs in the HUR-bound transcriptome (±25 nt around the peak summit), with two replicates for each species. (**D**) The line plot (smoothing bin width of 3 nt) displays the frequency of UUUs ± 100 nt at HUR peak summits (color-coded by species) and control sites residing 1000 nt distant from the peak summit (gray, color-coded by species), $n = 2$. (**E**) Binding site enrichments (average fold changes, *y*-axis) are shown as a function of the frequency of UUUs (*x*-axis) for all HUR orthologs. LOESS-fitted smoothed regression lines with the 95% confidence interval display the averaged enrichments (Supplementary Fig. S4B) across all orthologs (left). Box plots (right) denote enrichment distributions (ortholog average fold change) of each UUU bin, $n = 2$. (**F**) Box plots demonstrate peak fold changes per ortholog. Asterisks (top) represent statistical significance (two-tailed Wilcoxon rank sum test, $***P < 0.001$). The numbers of HUR-binding sites per species are indicated (bottom), $n = 2$. (G–I) Plots represent (**G**) average HUR-binding (scored as $\log_2$-transformed average binding scores, subsampled across six levels of orthologous species intersections), (**H**) degree of sequence conservation of HUR-bound sites (dark blue) compared with adjacent unbound control sites (light blue) (according to 100 vertebrate phyloP scores, subsampled), and (**I**) proportional frequency (ranging from 1 to 5) of UUU motifs within HUR-bound sites shown for intersections for each orthologous species (1–6, representing increasing evolutionary conservation), $n = 2$. (**J**) The box plot elucidates peak fold changes over all binding sites of zebrafish ($n = 1903$) and human ($n = 2069$) HUR using zebrafish RNA. Asterisks represent statistical significance (two-tailed Wilcoxon rank sum test, $***P < 0.001$), $n = 2$. (**K**) Pie chart shows the percentage of HUR-binding sites for a given RNA feature (blue, color-coded) in the combinatorial human and zebrafish cross-species RAPseq assays, $n = 2$. (**L**) The proposed model summarizes the co-evolutionary dependencies between HUR protein and transcriptomes in vertebrates. Top: from zebrafish to human, HUR-binding transitioned from mRNA processing to mRNA stability (inverse triangles). Middle: during evolution, the size of the actively transcribed genome (light blue line) doubled from zebrafish to human while HUR-binding sites increased (dark blue line). Bottom: HUR-binding specificity for UUUs remained unchanged (motif in the background) but HUR-binding affinities changed across vertebrates.

ing events (Fig. 4C). UUUs were centrally positioned within the peak summit over background control sites (Fig. 4D) (see Materials and methods) and fold changes were proportional to the number of triplets per binding site (Fig. 4E). This implies conserved substrate recognition and cooperative binding by the RRMs. However, frog and zebrafish HURs showed significantly lower enrichments compared with the other ortholog HURs (two-tailed Wilcoxon rank sum test, $P < 0.001$) (Fig. 4F), indicating differences in RNA-binding affinities due to amino acid changes.

We assessed the conservation of HUR-binding sites by averaging peak binding score values across tetrapod orthologs and clustering RNA-binding events by species overlap (Supplementary Fig. S4B). Shared sites among multiple species had significantly higher peak binding scores than species-specific RNA-binding events (Fig. 4G). To assess the degree of evolutionary conservation of HUR-binding sites, we used phyloP scores obtained from 100 vertebrate genome alignments. We found that both shared and unique HUR-binding sites were more constrained than adjacent regions used as control (Fig. 4H) (see Materials and methods). Remarkably, shared RNA-binding sites were under stronger purifying selection than those unique to each ortholog (Fig. 4H). This pattern mirrored the increase in average peak binding scores (Fig. 4G) and UUU motif frequencies (Fig. 4I), indicating positive selection for UUU sequences during vertebrate evolution to support HUR-binding.

Given that zebrafish HUR is most divergent across the vertebrate lineage, we profiled zebrafish and human HUR proteins on zebrafish RNA (Fig. 4B). We found that both HUR orthologs recognized UUUs in zebrafish RNA with similar specificity and cooperativity (Supplementary Fig. S4C, D), but zebrafish HUR exhibited significantly lower enrichments than human HUR (Fig. 4J). By inspecting HUR-binding events across the transcript body, we noticed a higher proportion of intronic binding sites in the zebrafish transcriptome, whereas human HUR preferentially bound 3′UTRs (Fig. 4K). To explore functional consequences of this divergence, we analyzed 59 GO terms commonly bound in human and zebrafish transcriptomes. In this shared gene set, zebrafish HUR bound nearly equally to zebrafish introns (45%) and 3′UTRs (55%), while human HUR preferentially bound human 3′ UTRs (86%) over introns (14%) (Supplementary Fig. S4E). We next inspected the proportion of HUR-binding sites within introns and 3′UTRs for each of the 59 GO terms (Supplementary Fig. S4F) and discovered the highest divergence in genes involved in monosaccharide metabolism, in which >60% of zebrafish HUR-binding sites were intronic, compared with 15% in human (Supplementary Fig. S4G). Given that HUR is the only paralog with both nuclear and cytoplasmic functions in vertebrates and that ELAV proteins in invertebrates are strictly nuclear [53], our results suggest an evolutionary shift in HUR function.

In conclusion, our cross-species combinatorial RAPseq experiments demonstrated that HUR-binding to UUUs is conserved across vertebrates, with selective expansion of vertebrate HUR-binding sites and increasing affinities from zebrafish to human. This functional transition from pre-mRNA processing in zebrafish to mRNA stability regulation in human (Fig. 4L) suggests an evolutionary shift in HUR function.

Cancer-associated variants of the IGF2BP family exhibit altered RNA-binding activity

To investigate the impact of cancer-associated mutations on RNA-binding activity, we examined recurrent somatic mutations in IGF2BP genes across different types of cancer. The IGF2BP family consists of three paralogs (IGF2BP1, IGF2BP2, and IGF2BP3). These paralogs play roles in regulating metabolism, cell growth and differentiation, as well as stress responses by stabilizing and enhancing the translation of mRNAs that encode key metabolic enzymes and growth-promoting factors [54]. In both human and mouse, IGF2BP genes exhibit oncofetal expression patterns with high expression levels during embryonic development and in liver cancer [55, 56] (Supplementary Fig. S5A–C). We identified one frequently mutated amino acid in IGF2BP1 (R167C/H) and two in IGF2BP3 (I474M and R525C) (Fig. 5A). While frequent missense mutations were not present in IGF2BP2, a cancer-associated splicing variant (isoform B) had been previously identified [57].

To evaluate the functional consequences, we generated wild-type IGF2BP paralogs and cancer-associated variants, and analyzed their RNA-binding capacities using RAPseq (Fig. 5B). The paralogs exhibited highly similar binding preferences towards CA-rich motifs, consistent with previous reports [58, 59]. We observed that these binding preferences were maintained in the cancer-associated variants (Spearman's rank correlation coefficient, median = 0.965, range: 0.92–0.99, $n = 28$) (Supplementary Fig. S5D). In contrast, k-mer enrichment profiles differed when comparing IGF2BP paralogs and variants with unrelated RBPs, such as RBFOX2 or YTHDF1 (Supplementary Fig. S5D).

Upon examining RNA-binding strength, we noted that IGF2BP1-R167C exhibited binding comparable to the wild-type, while the R167H variant showed weaker signal intensities (Supplementary Fig. S5E). IGF2BP2 isoform A showed slightly stronger binding to RNA targets compared to IGF2BP2 isoform B (Fig. 5C). The IGF2BP3 variants (I474M, R525C) displayed reduced binding strength compared to the wild-type (Supplementary Fig. S5E). Differential RNA-binding analysis revealed that the IGF2BP1-R167C variant barely differed from the wild-type. In contrast, the IGF2BP1-R167H variant lost 493 binding sites and gained only 15, with altered transcript feature distribution in the exon and 3′UTR, and a loss of RNA targets associated with platelet regulation (Supplementary Fig. S5F–H). The two IGF2BP2 variants showed a small and balanced number of differently bound sites (41 for IGF2BP2 isoform B and 68 for IGF2BP2 isoform A) (Fig. 5D). Both IGF2BP3 variants bound to fewer sites than the wild-type (2 versus 97 for IGF2BP3-I474M and 12 versus 99 for IGF2BP3-R525C) (Supplementary Fig. S5F). The IGF2BP3-I474M variant exhibited increased binding within coding sequences and decreased association with transcripts related to mitochondrial activity (Supplementary Fig. S5G, H). The IGF2BP3-R525C variant showed loss of binding to RNA targets associated with bacterial response pathways (Supplementary Fig. S5H).

We observed that the differentially bound sites of IGF2BP2 isoform B were more frequent in exons compared with those of IGF2BP2 isoform A, which were more enriched in coding sequences and 5′UTRs (χ^2 test, $P < 0.001$) (Fig. 5E). Among differentially bound exons, IGF2BP2 isoform B preferentially

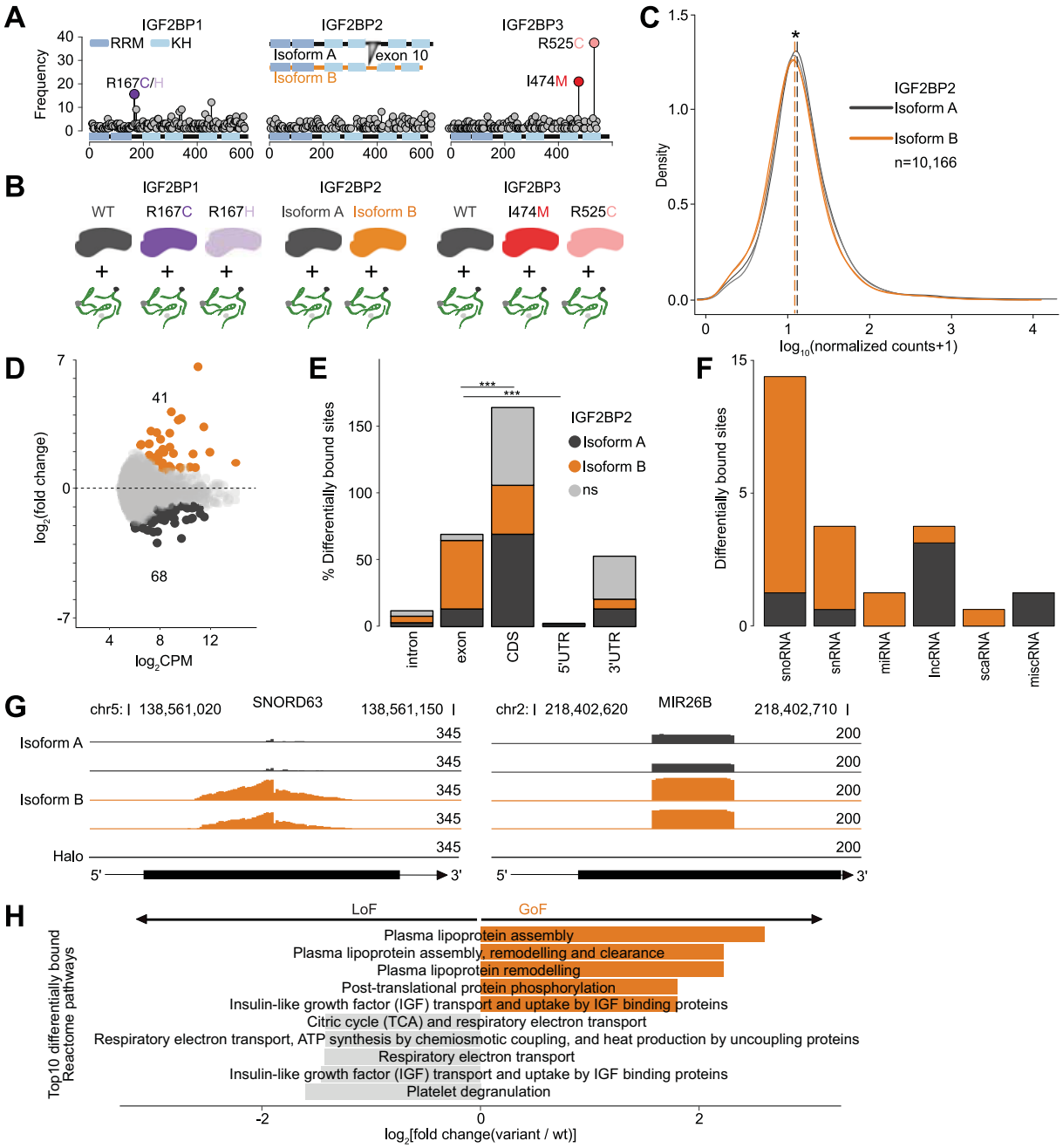


Figure 5. Pathological variants of the IGF2BP family cause functional gains and losses in RNA-binding activity. (**A**) Schematic diagrams illustrate the domain organization (RRM and KH domains) of each IGF2BP paralog (*x*-axis: length in amino acids). Lollipops show the frequency of amino acid mutation (*y*-axis) in cancer patients. Mutants selected for RAPseq profiling are color-coded. The pathogenic IGF2BP2 isoform B, which results from skipping exon 10, is shown. (**B**) Schematic representation of RAPseq assays performed to profile IGF2BP paralogs and pathological variants (color-coded). (**C**) Density plots represent the distribution of RAPseq signals (read counts normalized to HaloTag control in $\log_{10}$) for each IGF2BP2 isoform. Dashed lines indicate medians. The number of observations (*n*) is shown. An asterisk represents statistical significance of normalized counts (unpaired two-sample *t*-test, $*P < 0.05$), $n = 2$. (**D**) The MA plot demonstrates differentially bound sites (DBS) between IGF2BP2 isoform A and isoform B. The *y*-axis shows the $\log_2$ fold changes, and the *x*-axis shows the average signal per binding site (in $\log_2$ counts per million). Sites significantly enriched in isoform A (dark gray), isoform B (orange), or unchanged (light gray) are indicated (FC >2, FDR-adjusted $P < 0.05$), $n = 2$. (**E**) The stacked bar plot shows the distribution of sites across major RNA features that are either uniquely bound by IGF2BP2 isoform A (dark gray) and isoform B (orange) or by both (light gray). Bars of the same color sum up to 100%. Asterisks represent *P*-values (χ^2 test with continuity correction, $***P < 0.001$) for statistically significant differences of bound sites in exons compared with coding sequences and 5′UTR for IGF2BP2 isoform B, $n = 2$. (**F**) The stacked bar plot depicts the number (*y*-axis) and type (*x*-axis) of differentially bound ncRNA biotypes by IGF2BP2 isoform A (dark gray) and isoform B (orange), $n = 2$. (**G**) Genome browser tracks illustrate binding profiles of IGF2BP2 isoform A (dark gray) and isoform B (orange) relative to HaloTag control (black) at *SNORD63* and *miR-26B* loci. Tracks indicate genomic coordinates (*x*-axis, in base pairs) and normalized read density (RPM, *y*-axis), $n = 2$. (**H**) Horizontal bar plots display the top 10 Reactome pathways differentially bound between IGF2BP2 isoform A and IGF2BP2 isoform B. Pathways with gain of function (GoF, orange) or loss of function (LoF, gray) in IGF2BP2 isoform B compared with IGF2BP2 isoform A are shown (FC > 2, FDR-adjusted $P < 0.05$), n = 2.

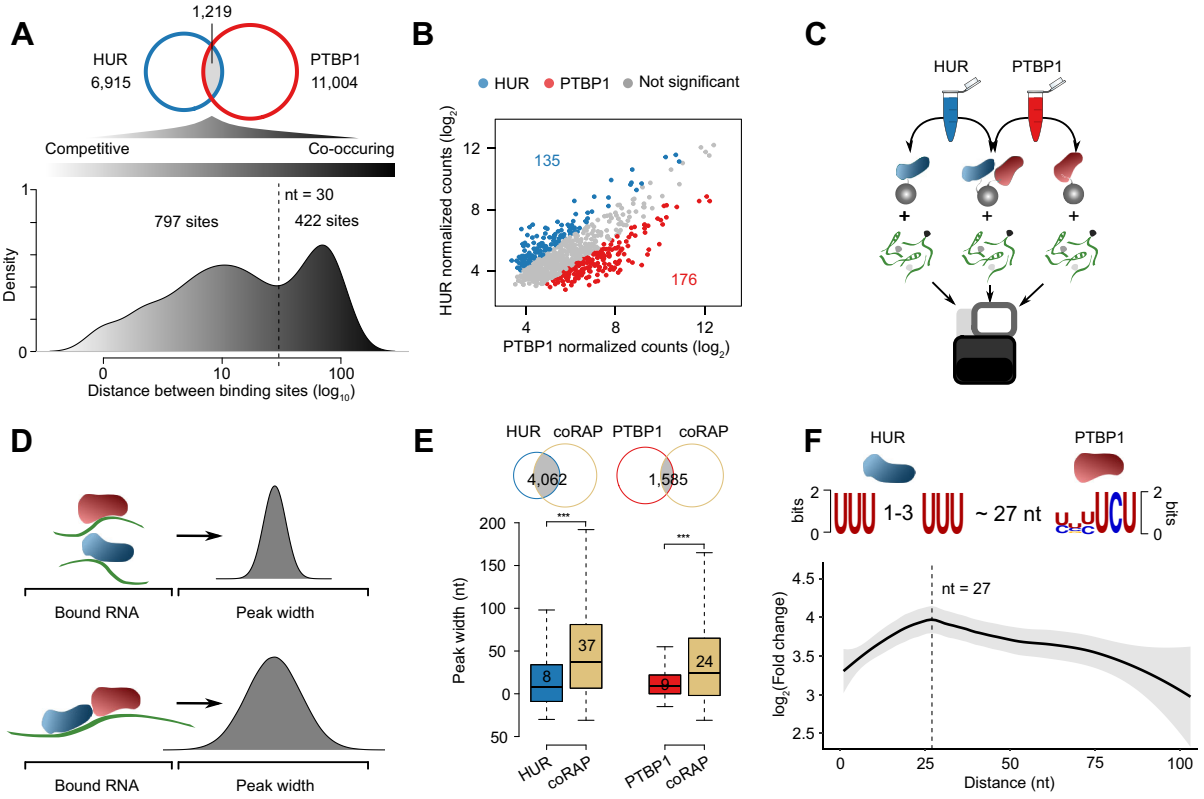


Figure 6. HUR and PTBP1 control RNA processing cooperatively with an optimal distance between their binding sites. (**A**) Venn diagram shows the overlap between HUR and PTBP1 peaks (top). The density plot (bottom) demonstrates the distance between peak summits of overlapping peaks. The *x*-axis represents the distance in nucleotides ($\log_{10}$ scale, band width 0.15). The bimodal distribution is separated at 30 nt (dashed line), $n = 2$. (**B**) The scatter plot compares HUR- and PTBP1-bound competitive peaks (dots, read count normalized, $\log_2$-transformed). Peaks either differentially bound by HUR (blue), PTBP1 (red), or not differentially bound (gray) are highlighted (FC > 2, FDR-adjusted $P < 0.05$), $n = 2$. (**C**) Schematic representation of the co-RAPseq assay for HUR (blue) and PTBP1 (red) using the same RNA substrate (green). (**D**) Illustration of the expected effect on peak widths of cooperative binding events. (**E**) Venn diagrams (top) show the overlap of binding peaks identified in individual RAPseq assays for HUR (blue, left) and PTBP1 (red, right), respectively, with those detected in co-RAPseq (yellow). The intersections represent the binding sites used to assess the differences in peak widths. Box plots (bottom) demonstrate differences in peak widths between the single RAPseq and co-RAPseq assays. The *y*-axis shows peak deviation from the median library size in nucleotides per assay. Asterisks represent statistical significance of two-sided Wilcoxon rank sum tests between single RAPseq and co-RAPseq assays (***$P < 0.001$), $n = 2$. (**F**) Model summarizing the cooperative binding of HUR (blue) and PTBP1 (red) to their respective motifs (top). LOESS-fitted regression lines depict the trend in binding site enrichments ($\log_2$-transformed fold change, *y*-axis) relative to the distance between the HUR and PTBP1 motifs (*x*-axis). The fold change in RNA-binding capacity is highest at 27 nt (dashed line).

bound to ncRNAs ($n = 22$) compared to IGF2BP2 isoform A ($n = 10$), with the highest differences in small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs) (Fig. 5F, G). At the pathway level, we observed significant differences between the IGF2BP2 isoforms despite the small effect sizes (FDR-adjusted $P < 0.05$) (Fig. 5H). The cancer-associated IGF2BP2 isoform B showed increased and decreased binding to transcripts involved in lipoprotein assembly and mitochondrial respiratory electron transport, respectively. These findings suggest that even modest changes in isoform-specific binding can redirect IGF2BP2 function toward distinct cellular pathways.

In summary, our results demonstrate that cancer-associated IGF2BP variants impact RNA-binding capacities in paralog-specific ways. Variants of IGF2BP1 and IGF2BP3 primarily alter binding strength without changing target binding specificity, while the IGF2BP2 isoform B changes target selection without strongly affecting global binding affinity. These changes could disrupt post-transcriptional regulation by weakening RBP–RNA interactions or shifting target selection, potentially unbalancing the RNA pool and contributing to tumor progression.

HUR and PTBP1 cooperate post-transcriptionally by maintaining a specific distance between binding sites

Previous reports have shown that HUR and PTBP1 regulate mRNA processing through both cooperative and competitive mechanisms [60, 61]. To investigate their binding relationship, we examined the overlap of their binding sites in single RAPseq assays. We identified a bimodal distribution of inter-peak distances, separated at 30 nt. Proximal sites (0–30 nt, $n = 797$) likely reflect competitive binding events, while distal peaks (31–100 nt, $n = 422$) indicate cooperative binding of the two RBPs (Fig. 6A). We statistically inferred the outcome of competitive events and determined that 17% (135/797) and 22% (176/797) of the proximal peaks could be attributed to either HUR or PTBP1, respectively (fold change > 2, FDR-adjusted $P < 0.05$) (Fig. 6B).

To directly assess cooperative RNA-binding, we developed co-RAPseq, in which RNA-binding capacity is assayed in the simultaneous presence of both RBPs (Fig. 6C). We hypothesized that unlike single RBP assays, cooperative events determined by co-RAPseq results in wider binding sites (peak width) due to the concurrent presence of the two proteins on

the same RNA substrate (Fig. 6D). While the peak width of the single RBP assays was 8 nt (HUR) and 9 nt (PTBP1) above the median library size, co-RAPseq assays with both factors resulted in significantly wider peak distributions of 37 nt and 24 nt, respectively (two-tailed Wilcox rank sum test, $P < 0.001$) (Fig. 6E). By analyzing co-RAPseq enrichment scores in relation to the spacing between HUR and PTBP1 motifs, we found that cooperative binding was optimal at a 27 nt distance (Fig. 6F). This optimal distance was not observable either in the single HUR and PTBP1 RAPseq assays or when using unbound control sites (Supplementary Fig. S6A–C).

Across the single or co-RAPseq assays, we identified 6885 genes with at least one RBP-binding event. Of those, 873 were uniquely detected in co-RAPseq (Supplementary Fig. S6D), suggesting that under steady-state conditions the post-transcriptional regulation of these gene products may depend on the co-activity of HUR and PTBP1. Moreover, 86% (2874/3324) of HUR-bound and 59% (2855/4862) of PTBP1-bound genes overlapped with the co-RAPseq assay (Supplementary Fig. S6D). Of these genes, 19% (536/2874 for HUR) and 22% (636/2855 for PTBP1) were significantly stronger bound when both RBPs acted together (Supplementary Fig. S6E). Functional enrichment analysis revealed that genes either uniquely or differentially bound in co-RAPseq were significantly associated with innate immune response pathways (Supplementary Fig. S6F), consistent with previously reported roles for each factor individually [62–66]. Remarkably, the strongest cooperative binding events occurred within introns of these genes, suggesting that HUR and PTBP1 coordinate their activity during mRNA pre-processing [67–69] (Supplementary Fig. S6G).

mod-RAPseq reveals modification-dependent RNA-binding activities

RNA modifications represent a pervasive and dynamic layer of gene regulation. To assess the transcriptome-wide relevance of RNA modifications in RBP–RNA interactions, we developed mod-RAPseq. We generated a modification-devoid RNA substrate by *in vitro* transcribing the native RNA substrate using T7 RNA polymerase (Fig. 7A). When comparing modification-containing native substrates used in RAPseq versus T7-transcribed unmodified substrates used in mod-RAPseq, we found nearly identical transcript distributions and high correlations (Spearman's rank $\rho = 0.93$ and Pearson correlation coefficient $r = 0.97$), confirming that *in vitro* transcription maintains the global RNA composition (Fig. 7B).

^{6}A is the most abundant mRNA modification in eukaryotes with critical roles in development, stress responses, and diseases [46]. We applied mod-RAPseq to determine modification-dependent targets of the m^{6}A binder YTHDF1 [46]. According to previous reports [46], YTHDF1 binds to a preferred GRAC (R is an A or G) motif of the m^{6}A DRACH consensus sequence (D is A, U, or G, and H is A, U, or C). About 81% (965/1 189) of the identified YTHDF1-binding sites were strictly dependent on the presence of m^{6}A (Fig. 7C) and the binding affinity decreased significantly in the absence of RNA modifications (two-tailed Wilcox rank sum test, $P < 0.001$) (Fig. 7D), corroborating m^{6}A dependency in RNA-binding. We observed that the 5-mer GGACU was the most abundant m^{6}A-DRACH motif (Fig. 7E), with 4-mer submotifs GGAC and GACU contributing to weaker but still frequent interactions (Fig. 7F–H). Binding sites with the pentamer motif exhibited higher enrichment scores than sites containing only tetramers (Fig. 7F–H), supporting that GGACU is the optimal binding motif, and the two suboptimal tetramers (GGAC and/or GACU) can increase the frequency of YTHDF1 interactions with RNA. Sites with m^{6}A-modified GGACU motifs showed stronger binding than their unmodified counterparts (two-tailed Wilcox rank sum test, $P < 0.01$) (Fig. 7F–H), confirming m^{6}A dependency of YTHDF1 [46].

We next investigated pathways that could be affected by the presence or absence of YTHDF1 m^{6}A-binding. We found 56 significantly enriched pathways (FDR-adjusted $P < 0.05$), with the highest m^{6}A dependency in genes involved in hormone transport and stem cell population maintenance (Fig. 7I). We confirmed the presence of m^{6}A in the top four binding sites in *SS18*, *KDM2B*, *NIBAN2*, and *LIF* mRNAs using public m^{6}A-seq data [6] (Fig. 7J).

To test whether mod-RAPseq can identify other RNA modification-dependent RBP-binding events, beyond m^{6}A, we applied mod-RAPseq to YBX1, a known reader of 5-methylcytosine (m^{5}C) [70]. m^{5}C is less abundant than m^{6}A but is gaining more attention due to its potential functional roles. Our analysis revealed that 83% (2035/2447) of the identified YBX1-binding sites depend on the presence of RNA modifications (Supplementary Fig. S7A), and their absence significantly reduced binding affinity (two-tailed Wilcoxon rank-sum test, $P < 0.001$) (Supplementary Fig. S7B). While earlier reports [41, 70] identified pyrimidine-rich sequences as preferred YBX1-binding motifs, our analysis additionally revealed a strong enrichment for CGA-containing motifs (Supplementary Fig. S7C). The top 10 CGA-containing motifs exhibited higher enrichment scores compared with motifs depleted of RNA modifications (Supplementary Fig. S7D). Notably, 32% (457/1437) of YBX1 m^{5}C-dependent targets overlapped with known m^{5}C methylated genes in T24 bladder cancer cells [50] (Supplementary Fig. S7E).

In conclusion, mod-RAPseq enables precise, transcriptome-wide identification of RNA modification-dependent RBP–RNA interactions. By contrasting binding profiles with and without modifications, this approach uncovers distinct roles for modifications, such as m^{6}A and m^{5}C, in shaping RBP-binding activities.

Discussion

Our study establishes RAPseq as a versatile and quantitative method for profiling RBP–RNA interactions *in vitro* using recombinant RBPs and native RNA substrates that retain intrinsic modifications. RAPseq captures sequence-, structure-, and modification-dependent RBP–RNA interactions with high quantitative accuracy. Validation with well-characterized RBPs, including HUR, PTBP1, and YTHDF1, confirmed its ability to recapitulate known binding motifs, such as UUUs, polypyrimidine tracts, and DRACH motifs, respectively, underscoring its reliability across diverse binding modalities. Compared with established methods such as RBNS, RNAcompete, and eCLIP, RAPseq demonstrated superior resolution in binding motif discovery.

RAPseq has enabled several advanced applications. First, we demonstrate the scalability of RAPseq by profiling 37 RBPs, including 11 cRBP and 26 ncRBPs. Our analysis revealed that ncRBPs exhibit specialized, rather than global, post-transcriptional regulatory roles. Compared with cRBPs, ncRBP–RNA interactions were characterized by fewer and

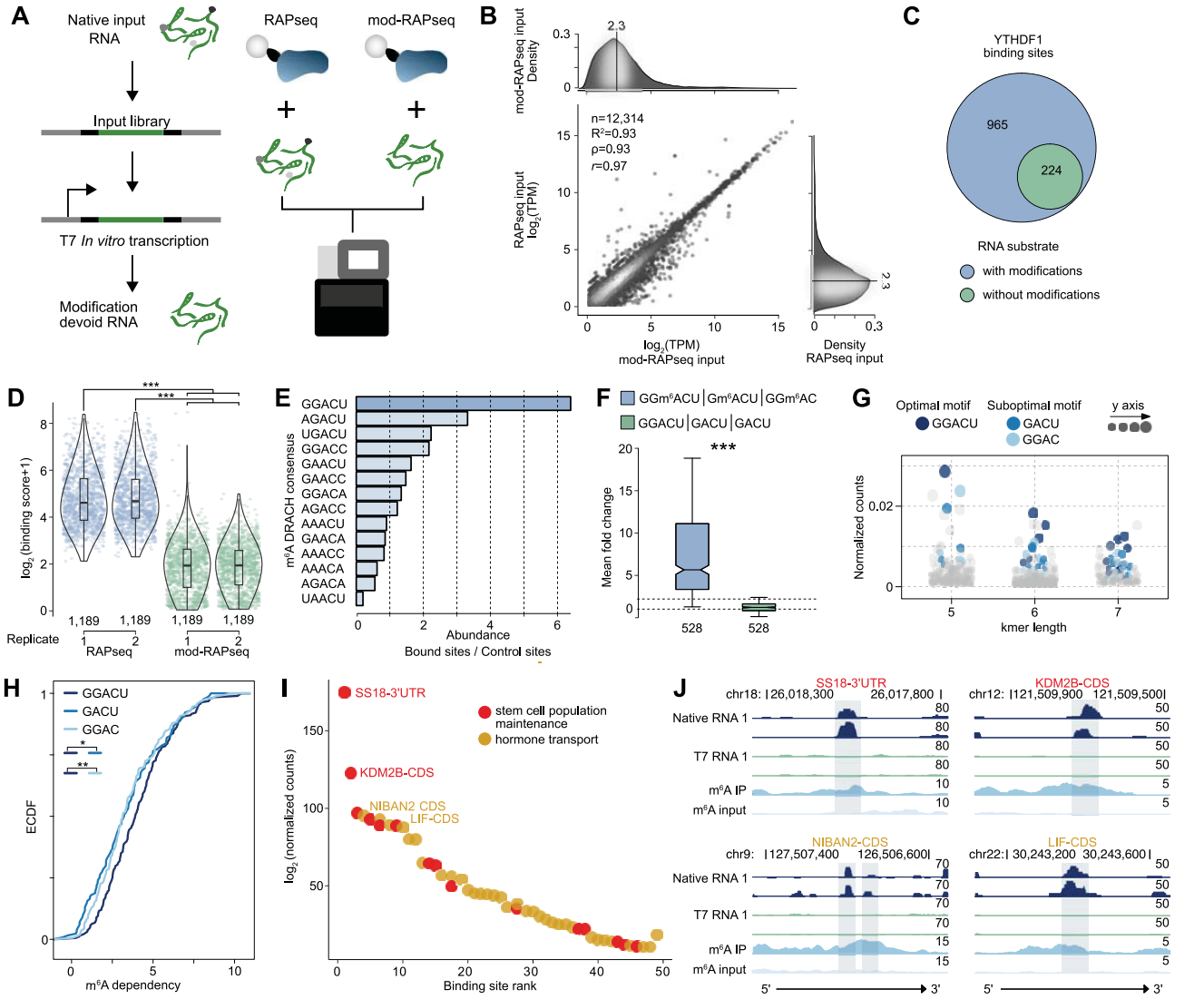


Figure 7. mod-RAPseq identifies m^{6}A- and m^{5}C-dependent RNA recognition. (**A**) Schematic illustration of the mod-RAPseq method used to assess RNA modification-dependent RBP-binding. (**B**) The scatter plot compares RNA profiles between mod-RAPseq (*in vitro* transcribed RNA) and RAPseq (native RNA) substrates ($\log_2$-transformed transcripts per million, TPM). The plot includes the number of observations, Spearman's rank (ρ) or Pearson (r) correlation coefficient, and adjusted R^2 of linear regression. Density plots (right and top, bandwidth 0.2) show the overall distributions of TPMs and their medians (2.3 $\log_2$ TPM, dashed line). (**C**) Venn diagram overlaps YTHDF1-binding sites identified in native RAP and the mod-RAP substrates, $n = 2$. (**D**) Violin plots with overlaid box plots compare YTHDF1 enrichments ($\log_2$-transformed peak binding scores + 1) across replicates in RAPseq (blue points) and mod-RAPseq (green points). Each dot represents one binding site (n = 1189). Asterisks represent statistical significance (two-tailed Wilcoxon rank sum test ***$P < 0.001$). (**E**) The bar plot shows the frequency of DRACH 5-mer motifs within YTHDF1-binding peaks normalized by their abundance at control sites. The top motif is highlighted (dark blue), $n = 2$. (**F**) The box plot compares YTHDF1 fold change enrichments at binding sites containing the GGACU motif in RAPseq (blue) versus mod-RAPseq (green) substrates ($n = 528$). Dashed lines mark fold change thresholds at 1 and 2. Asterisks represent statistical significance (two-tailed Wilcoxon rank sum test ***$P < 0.001$), $n = 2$. (**G**) The dot plot highlights the enrichment of k-mers of various lengths containing the GGACU motif or its two encompassed 4-mers, GGAC and GACU. The fraction of k-mer counts corresponds to weighted enrichment ratios between bound versus control sites (*y*-axis), $n = 2$. (**H**) The line plot of empirical cumulative distribution functions (ECDFS) shows YTHDF1-binding dependency on m^{6}A presence in each motif at each binding site. The dependency is computed as the $\log_2$ ratio of peak binding scores between native RNA or *in vitro* transcribed RNA. Asterisks represent statistical significance (two-tailed Wilcoxon rank sum test, *$P < 0.05$, **$P < 0.01$), $n = 2$. (**I**) The dot plot ranks m^{6}A-dependent YTHDF1-binding sites within mRNA of genes from two enriched GO terms. The RAPseq binding site signal (library size $\log_2$-transformed normalized read counts, *y*-axis) is ranked (*x*-axis). The top four target genes are highlighted, $n = 2$. (**J**) Genome tracks show YTHDF1-binding to m^{6}A-modified regions within the 3′UTRs and coding sequences of the top two binding sites of the two GO terms (encompassing *SS18*, *KDM2B*, *NIBAN2*, and *LIF*) (I). Genomic locations with a scale bar indicate the length of the genomic region in bases. Tracks show normalized read densities (RPM) (*y*-axis) for two replicates using native (dark blue) or T7 *in vitro* transcribed (light blue) input RNA as well as m^{6}A-specific RNA immunoprecipitation (IP, blue), and the respective input control are presented. Vertical blue lines highlight bound sites. Arrows indicate the direction of transcription.

more selective binding events, often focused on specific transcript regions. Some RBPs co-regulated shared target gene networks, suggesting functional convergence with known regulatory pathways. For example, STMN1 co-regulated gene networks enriched for functions already associated with their canonical, non-RNA-binding roles, suggesting potential moonlighting roles in gene regulation [52, 71]. Other affinity tag-based methods, including cross-linking and analysis of cDNAs (CRAC) [72] or cross-linking, ligation and sequencing of hybrids (CLASH) [73] could be applied to ncRBPs, but both methods require prior genetic manipulation of cells, efficient cross-linking, and cellular context, which may limit scalability for systematic interrogation of large sets of ncRBPs.

Second, cross-species profiling revealed that HUR-binding to UUUs is conserved across vertebrates, with species-specific preferences for intronic versus 3′UTR-binding sites, suggesting evolutionary divergence in RNA processing. This is consistent with transcriptomic shifts observed in vertebrate evolution [56, 24]. While cross-species analysis can also be performed with eCLIP, it depends on the availability of high-quality and cross-reactive antibodies that recognize the RBP orthologs in different species. RAPseq circumvents this limitation by using recombinant proteins.

Third, we applied RAPseq to characterize cancer-associated IGF2BP variants. IGF2BP1 and IGF2BP3 mutants retained motif specificity but showed reduced binding strength. In contrast, the cancer-associated IGF2BP2 isoform B exhibited altered target selection, particularly among short ncRNAs and metabolic transcripts, implying a possible role in metabolic reprogramming during tumorigenesis [74, 75]. While other *in cellulo* approaches can assess individual RBP variants generated by gene editing, RAPseq allows direct, high-throughput comparison of binding behaviors across isoforms and mutants in a highly adaptable and efficient manner.

Fourth, our co-RBP assay revealed that HUR and PTBP1 cooperate post-transcriptionally, with optimal regulatory potential depending on spatial proximity between binding sites. This supports the emerging model that RBPs exert concerted functions through spatially coordinated interactions [36, 76]. While *in vivo* approaches such as TRIBE and STAMP can also detect RBP–RBP cooperativity by fusing RNA editing enzymes to RBPs and identifying co-editing events, their effectiveness depends on enzyme co-expression, fusion protein activity, and background editing levels [77]. In contrast, our *in vitro* approach enables direct, controlled assessment of cooperative binding, offering mechanistic insights that complement but are not confounded by cellular complexity.

Fifth, we developed mod-RAPseq, a modification-depleted variant of the native RAPseq substrate, to specifically investigate modification-dependent RBP–RNA interactions. Using this approach, we confirmed m^6A- and m^5C-dependent binding of YTHDF1 and YBX1, respectively, consistent with prior reports showing their reliance on RNA modifications for recognition and regulatory function [45, 70]. Conventional methods, such as eCLIP, combined with genetic knockout or knockdown of individual writer enzymes can reveal modification dependence, but they are time-consuming, typically limited to one modification at a time, and require efficient genetic manipulation. In contrast, mod-RAPseq offers a streamlined alternative that enables parallel assessment of multiple modification-dependent interactions without genetic manipulation.

RAPseq is uniquely suited to resolve dynamic regulatory features across canonical and non-canonical RBPs, species evolution, and healthy and disease states. Future improvements, such as enriching nuclear RNA to better capture currently under-represented intronic sequences (Supplementary Fig. S2G) could expand the utility of RAPseq to study alternative splicing and co-regulatory RBP dependencies [68]. The growing number of ncRBPs without known RNA-binding functions, but with well-established roles in cellular metabolism, poses important questions regarding the contextuality of their RNA-binding activities. It remains to be explored whether RNA binding, enzymatic, and metabolic functions occur simultaneously or are mutually exclusive. RAPseq could be employed to investigate RNA–enzyme–metabolite networks [78] by probing RBP–RNA interactions under varying metabolic conditions and across distinct transcriptomes. This may reveal how these networks interact dynamically and shed light on context-specific regulatory mechanisms.

Collectively, our findings position RAPseq as a flexible and scalable method facilitating consortia-level studies within individual laboratories. This tool enables acceleration of the functional characterization of the RBPome by decoding the post-transcriptional regulatory code and advancing research that would otherwise be unfeasible.

Limitations

RAPseq enables the sensitive detection of RNA–protein interactions in a controlled *in vitro* setting, uncovering binding events that may be hidden *in vivo* due to RNA structural constraints. However, context-dependent RBP–RNA interactions that rely on post-translational modifications, chaperone-mediated multi-subunit assembly, and dynamic subcellular localization may not be captured *in vitro*. Additionally, the use of the HaloTag fusion in RAPseq could potentially impact the three-dimensional conformation and activity of the RBP, a limitation shared with other fusion-based methods like TRIBE and STAMP. Since RAPseq focuses on identifying the binding potential of an RBP rather than its physiological binding events, integrating RAPseq with cellular assays that maintain native interactions can help differentiate biochemical binding capacity from biologically relevant interactions.

The use of chemically fragmented RNA (∼30 nt) enhances substrate accessibility and simplifies library preparation, but it may disrupt RNA structures that rely on longer nucleotide stretches, potentially removing binding sites that require extended sequence context or cooperative interactions. Nonetheless, our profiling of the RBP IRP1, which binds iron-response elements, forming defined stem–loop structures indicates that RAPseq can detect interactions within short-range structural conformations.

Future refinements, such as enriching nuclear RNA, could improve the capture of currently under-represented intronic sequences (Supplementary Fig. S2G) and expand the application of RAPseq to study alternative splicing and co-regulatory RBP dependencies [68]. While RAPseq libraries contain a significant number of reads mapping to rRNA and tRNA genes (Supplementary Table S3), which can limit sequencing depths for other RNA biotypes, these reads

can provide valuable information on potential interactions with structured ncRNAs. If these interactions are not of interest, an rRNA depletion step could further enhance specificity.

Despite these considerations, RAPseq successfully identifies bona fide binding sites across 37 profiled RBPs with diverse protein lengths, domain architectures, families, and levels of intrinsic disorder, underscoring the method's robustness and versatility.

Acknowledgements

We would like to thank the National Genomics Infrastructure in Stockholm, the members of the Friedländer, Pelechano, and Kutter laboratories, as well as Carsten Daub for helpful feedback.

Author contributions: Riccardo Mosca (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), Carlos J. Gallardo-Dodd (Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), Qun Li (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), Christian Sommerauer (Data curation [equal], Formal analysis [equal], Methodology [equal], Writing – review & editing [equal]), Justas Šidiškis (Data curation [equal], Formal analysis [equal], Methodology [equal], Writing – review & editing [equal]), Jonas N. Søndergaard (Data curation [equal], Formal analysis [equal], Methodology [equal], Writing – review & editing [equal]), Claudia Kutter (Conceptualization [lead], Funding acquisition [lead], Methodology [equal], Project administration [lead], Resources [lead], Supervision [lead], Writing – original draft [lead], Writing – review & editing [lead]).

Supplementary data

Supplementary data are available at NAR online.

Conflict of interest

The authors declare no competing interests.

Funding

Knut & Alice Wallenberg foundation [KAW 2016.0174 to C.K.]; Ruth & Richard Julin foundation [2020-00294, 2022-00283] and 2023-00162 to C.K.]; SFO-SciLifeLab fellowship [SFO_004 to C.K.]; Swedish Research Council [2019-05165 and 2023-02780 to C.K.]; Cancerfonden [22 2246 Pj to C.K.]; KI-KID [2018-00904 and 2021-00308 to C.K.]; Robert Lundberg's Memorial Foundation [2022-01158 to C.S. and 2025-02130 to C.J.G.-D.]; ERASMUS+ [2021/092 to R.M.]; and the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at PDC [storage: 2024/22-144, 2024/23-563; compute: 2024/23-84, 2024/22-1263].

Data availability

Raw and processed sequencing data generated in this study have been submitted to ArrayExpress (E-MTAB-10834). All steps required for preprocessing RAPseq reads were automated using a Snakemake workflow available on Zenodo (https://doi.org/10.5281/zenodo.18316842), together with associated R scripts and software version information. Detailed descriptions of data filtering, visualization, and metafile generation are also provided at the same repository. Publicly available eCLIP-seq and RBNS data were retrieved from the ENCODE Project with the following accession numbers: HNRNPA1 (ENCFF797GSK), HNRNPC (ENCFF440ROZ), PTBP1 (ENCFF726SQU), RBFOX2 (ENCFF871NYM), YBX3 (ENCFF185OEI), HNRNPC (ENCSR569UIU), and RBFOX2 (ENCSR441HLP).

References

1. Perez-Perri JI, Noerenberg M, Kamel W *et al.* Global analysis of RNA-binding protein dynamics by comparative and enhanced RNA interactome capture. *Nat Protoc* 2021;16:27–60. https://doi.org/10.1038/s41596-020-00404-1
2. Queiroz RML, Smith T, Villanueva E *et al.* Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol* 2019;37:169–78. https://doi.org/10.1038/s41587-018-0001-2
3. Trendel J, Schwarzl T, Horos R *et al.* The human RNA-binding proteome and its dynamics during translational arrest. *Cell* 2019;176:391–403. https://doi.org/10.1016/j.cell.2018.11.004
4. Urdaneta EC, Vieira-Vieira CH, Hick T *et al.* Purification of cross-linked RNA–protein complexes by phenol–toluol extraction. *Nat Commun* 2019;10:990. https://doi.org/10.1038/s41467-019-08942-3
5. Alarcón CR, Lee H, Goodarzi H *et al.* N6-methyladenosine marks primary microRNAs for processing. *Nature* 2015;519:482–5. https://doi.org/10.1038/nature14281
6. Dominissini D, Moshitch-Moshkovitz S, Schwartz S *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012;485:201–6. https://doi.org/10.1038/nature11112
7. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 2016;530:441–6. https://doi.org/10.1038/nature16998
8. Huang H, Weng H, Sun W *et al.* Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* 2018;20:285–95. https://doi.org/10.1038/s41556-018-0045-z
9. Meyer KD. DART-seq: an antibody-free method for global m6A detection. *Nat Methods* 2019;16:1275–80. https://doi.org/10.1038/s41592-019-0570-0
10. Zhang L-S, Liu C, Ma H *et al.* Transcriptome-wide mapping of internal *N*7-methylguanosine methylome in mammalian mRNA. *Mol Cell* 2019;74:1304–16. https://doi.org/10.1016/j.molcel.2019.03.036.
11. Meyer KD, Saletore Y, Zumbo P *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 2012;149:1635–46. https://doi.org/10.1016/j.cell.2012.05.003
12. Squires JE, Patel HR, Nousch M *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* 2012;40:5023–33. https://doi.org/10.1093/nar/gks144
13. Ule J, Jensen KB, Ruggiu M *et al.* CLIP identifies nova-regulated RNA networks in the brain. *Science* 2003;302:1212–5. https://doi.org/10.1126/science.1090095
14. Licatalosi DD, Mele A, Fak JJ *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;456:464–9. https://doi.org/10.1038/nature07488

15. Hafner M, Landthaler M, Burger L *et al.* Transcriptome-wide identification of RNA-binding protein and MicroRNA target sites by PAR-CLIP. *Cell* 2010;141:129–41. https://doi.org/10.1016/j.cell.2010.03.009
16. König J, Zarnack K, Rot G *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 2010;17:909–15.https://doi.org/10.1038/nsmb.1838
17. Zarnegar BJ, Flynn RA, Shen Y *et al.* irCLIP platform for efficient characterization of protein–RNA interactions. *Nat Methods* 2016;13:489–92. https://doi.org/10.1038/nmeth.3840
18. Van Nostrand EL, Pratt GA, Shishkin AA *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 2016;13:508–14. https://doi.org/10.1038/nmeth.3810
19. Lorenz DA, Her H-L, Shen KA *et al.* Multiplexed transcriptome discovery of RNA-binding protein binding sites by antibody-barcode eCLIP. *Nat Methods* 2023;20:65–9. https://doi.org/10.1038/s41592-022-01708-8
20. Wheeler EC, Van Nostrand EL, Yeo GW. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip Rev RNA* 2018;9:e1436. https://doi.org/10.1002/wrna.1436
21. Lin C, Miles WO. Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions. *Nucleic Acids Res* 2019;47:5490–501. https://doi.org/10.1093/nar/gkz295
22. Ray D, Kazan H, Cook KB *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 2013;499:172–7. https://doi.org/10.1038/nature12311
23. Lambert N, Robertson A, Jangi M *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* 2014;54:887–900. https://doi.org/10.1016/j.molcel.2014.04.016
24. Kutter C, Brown GD, Gonçalves Â *et al.* Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet* 2011;43:948–55. https://doi.org/10.1038/ng.906
25. Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4
26. Smith TS, Heger A, Sudbery I. UMI-tools: modelling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res* 2017;27:491–9. https://doi.org/10.1101/gr.209601.116
27. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137
28. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033
29. Lawrence M, Huber W, Pagès H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118
30. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 2011;27:1653–9. https://doi.org/10.1093/bioinformatics/btr261
31. Chen Y, Chen L, Lun ATL *et al.* edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res* 2025;53:gkaf018. https://doi.org/10.1093/nar/gkaf018
32. Wu T, Hu E, Xu S *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;2:100141. https://doi.org.10.1016/j.xinn.2021.100141
33. Castello A, Fischer B, Eichelbaum K *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012;149:1393–406. https://doi.org/10.1016/j.cell.2012.04.031
34. Shannon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504. https://doi.org/10.1101/gr.1239303
35. Søndergaard JN, Sommerauer C, Atanasoai I *et al.* CCT3–LINC00326 axis regulates hepatocarcinogenic lipid metabolism. *Gut* 2022;71:2081–92. https://doi.org/10.1136/gutjnl-2021-325109
36. Van Nostrand EL, Freese P, Pratt GA *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* 2020;583:711–9. https://doi.org/10.1038/s41586-020-2077-3
37. Sayols S. rrvgo: a Bioconductor package for interpreting lists of gene ontology terms. *MicroPubl Biol* 2023. https://doi.org/10.17912/micropub.biology.000811
38. Vasaikar SV, Straub P, Wang J *et al.* LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res* 2018;46:D956–63. https://doi.org/10.1093/nar/gkx1090
39. Tang Z, Kang B, Li C *et al.* GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019;47:W556–60. https://doi.org/10.1093/nar/gkz430
40. Los GV, Encell LP, McDougall MG *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 2008;3:373–82. https://doi.org/10.1021/cb800025k
41. Ray D, Kazan H, Chan ET *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* 2009;27:667–70. https://doi.org/10.1038/nbt.1550
42. Lebedeva S, Jens M, Theil K *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* 2011;43:340–52. https://doi.org/10.1016/j.molcel.2011.06.008
43. Ripin N, Boudet J, Duszczyk MM *et al.* Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. *Proc Natl Acad Sci USA* 2019;116:2935–44. https://doi.org/10.1073/pnas.1808696116
44. Hentze MW, Caughman SW, Rouault TA *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 1987;238:1570–3. https://doi.org/10.1126/science.3685996
45. Xu C, Wang X, Liu K *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat Chem Biol* 2014;10:927–9. https://doi.org/10.1038/nchembio.1654
46. Wang X, Zhao BS, Roundtree IA *et al.* N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* 2015;161:1388–99. https://doi.org/10.1016/j.cell.2015.05.014
47. Begg BE, Jens M, Wang PY *et al.* Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nat Struct Mol Biol* 2020;27:901–12. https://doi.org/10.1038/s41594-020-0475-8
48. Feng H, Bao S, Rahman MA *et al.* Modeling RNA-binding protein specificity *in vivo* by precisely registering protein-RNA crosslink sites. *Mol Cell* 2019;74:1189–204. https://doi.org/10.1016/j.molcel.2019.02.002
49. Ilık İA, Glažar P, Tse K *et al.* Autonomous transposons tune their sequences to ensure somatic suppression. *Nature* 2024;626:1116–24. https://doi.org.10.1038/s41586-024-07081-0
50. Zarnack K, König J, Tajnik M *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 2013;152:453–66. https://doi.org/10.1016/j.cell.2012.12.023
51. Hafner M, Katsantoni M, Köster T *et al.* CLIP and complementary methods. *Nat Rev Methods Primers* 2021;1:1–23. https://doi.org/10.1038/s43586-021-00018-1
52. Gallardo-Dodd CJ, Kutter C. The regulatory landscape of interacting RNA and protein pools in cellular homeostasis and cancer. *Hum Genomics* 2024;18:109. https://doi.org/10.1186/s40246-024-00678-6
53. Good PJ, A conserved family of elav-like genes in vertebrates. *Proc Natl Acad Sci USA* 1995;92:4557–61. https://doi.org/10.1073/pnas.92.10.4557
54. Du Q-Y, Zhu Z-M, Pei D-S. The biological function of IGF2BPs and their role in tumorigenesis. *Invest New Drugs* 2021;39:1682–93. https://doi.org/10.1007/s10637-021-01148-9

55. Rudolph KLM, Schmitt BM, Villar D *et al.* Codon-driven translational efficiency is stable across diverse mammalian cell states. *PLoS Genet* 2016;12:e1006024. https://doi.org/10.1371/journal.pgen.1006024
56. Cardoso-Moreira M, Halbert J, Valloton D *et al.* Gene expression across mammalian organ development. *Nature* 2019;571:505–9. https://doi.org/10.1038/s41586-019-1338-5
57. Zhang J-Y, Chan EKL, Peng X-X *et al.* A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma. *J Exp Med* 1999;189:1101–10. https://doi.org/10.1084/jem.189.7.1101
58. Conway AE, Nostrand ELV, Pratt GA *et al.* Enhanced CLIP uncovers IMP protein–RNA targets in human pluripotent stem cells important for cell adhesion and survival. *Cell Rep* 2016;15:666–79. https://doi.org/10.1016/j.celrep.2016.03.052
59. Schneider T, Hung LH, Aziz M *et al.* Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat Commun* 2019;10:1–18. https://doi.org/10.1038/s41467-019-09769-8
60. Galbán S, Kuwano Y, Pullmann R, Jr *et al.* RNA-binding proteins HuR and PTB promote the translation of hypoxia-inducible factor 1α. *Mol Cell Biol* 2008;28:93–107. https://doi.org/10.1128/MCB.00973-07
61. Ji X, Liu Z, Gao J *et al.* N6-methyladenosine-modified lncRNA LINREP promotes glioblastoma progression by recruiting the PTBP1/HuR complex. *Cell Death Differ* 2023;30:54–68. https://doi.org/10.1038/s41418-022-01045-5
62. Christodoulou-Vafeiadou E, Ioakeimidis F, Andreadou M *et al.* Divergent innate and epithelial functions of the RNA-binding protein HuR in intestinal inflammation. *Front Immunol* 2018;9:2732. https://doi.org/10.3389/fimmu.2018.02732
63. Sueyoshi T, Kawasaki T, Kitai Y *et al.* Hu antigen R regulates antiviral innate immune responses through the stabilization of mRNA for polo-like kinase 2. *J Immunol* 2018;200:3814–24. https://doi.org/10.4049/jimmunol.1701282
64. Sasanuma H, Ozawa M, Yoshida N. RNA-binding protein Ptbp1 is essential for BCR-mediated antibody production. *Int Immunol* 2018;31:157–66. https://doi.org/10.1093/intimm/dxy077
65. Geng G, Xu C, Peng N *et al.* PTBP1 is necessary for dendritic cells to regulate T-cell homeostasis and antitumour immunity. *Immunology* 2021;163:74–85. https://doi.org/10.1111/imm.13304
66. Rothamel K, Arcos S, Kim B *et al.* ELAVL1 primarily couples mRNA stability with the 3′ UTRs of interferon-stimulated genes. *Cell Rep* 2021;35:109178. https://doi.org/10.1016/j.celrep.2021.109178
67. Izquierdo JM. Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition. *J Biol Chem* 2008;283:19077–84. https://doi.org/10.1074/jbc.M800017200
68. Tao Y, Zhang Q, Wang H *et al.* Alternative splicing and related RNA binding proteins in human health and disease. *Signal Transduct Target Ther* 2024;9:1–33.
69. Adamoski D, M dos Reis L, Mafra ACP *et al.* HuR controls glutaminase RNA metabolism. *Nat Commun* 2024;15:5620. https://doi.org/10.1038/s41467-024-49874-x
70. Chen X, Li A, Sun B-F *et al.* 5-Methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat Cell Biol* 2019;21:978–90. https://doi.org/10.1038/s41556-019-0361-y
71. Hentze MW, Castello A, Schwarzl T *et al.* A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 2018;19:327–41. https://doi.org/10.1038/nrm.2017.130
72. Granneman S, Kudla G, Petfalski E *et al.* Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci USA* 2009;106:9613–8. https://doi.org/10.1073/pnas.0901997106
73. Kudla G, Granneman S, Hahn D *et al.* Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc Natl Acad Sci USA* 2011;108:10010–5. https://doi.org/10.1073/pnas.1017386108
74. Le HTT, Sorrell AM, Siddle K. Two isoforms of the mRNA binding protein IGF2BP2 are generated by alternative translational initiation. *PLoS One* 2012;7:e33140. https://doi.org/10.1371/journal.pone.0033140
75. Nachmani D, Gutschner T, Reches A *et al.* RNA-binding proteins regulate the expression of the immune activating ligand MICB. *Nat Commun* 2014;5:4186. https://doi.org/10.1038/ncomms5186
76. Lang B, Yang J-S, Garriga-Canut M *et al.* Matrix-screening reveals a vast potential for direct protein–protein interactions among RNA binding proteins. *Nucleic Acids Res* 2021;49:6702–21. https://doi.org/10.1093/nar/gkab490
77. Owens MC, Liu KF. TRIBE-STAMP reveals new insights into the functions of RNA binding proteins. *Genes Dev* 2022;36:954–5.
78. Fradera-Sola A, Nischwitz E, Bayer ME *et al.* RNA-dependent interactome allows network-based assignment of RNA-binding protein function. *Nucleic Acids Res* 2023;51:5162–76. https://doi.org/10.1093/nar/gkad245

Received: June 27, 2025. **Revised:** December 18, 2025. **Accepted:** January 19, 2026