In Vivo Analysis of RNA Proximity Proteomes Using RiboPro

Xianzhi Lin[1 #], Kate Lawrenson[1, 2 #]

[1] Women's Cancer Program at Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

[2] Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

#Correspondence to: xianzhi.lin@cshs.org (XL) or kate.lawrenson@cshs.org (KL).

Functional and mechanistic annotation of uncharacterized long noncoding RNAs is challenging and often requires identification of interacting proteins. Here we report RiboPro, a flexible method that leverages the RNA-binding specificity of inactive Cas13b and proximity labeling activity of peroxidase APEX to permit rapid and unbiased discovery of target RNA binding proteins *in vivo*. RiboPro of poly(A)+ RNA reveals insights into poly (A)+ RNA nucleocytoplasmic transport, localization, turnover, and higher-order structure.

LncRNA transcripts do not encode protein, but they nonetheless play myriad roles in physiology and disease[1]. The spectrum of RNA binding proteins (RBPs) associated with a lncRNA transcript will largely dictate lncRNA activities[2]. RBPs control critical RNA activities including RNA editing, processing, stability, translation and transport[3]. Therefore, cataloguing the interacting proteins of lncRNAs is critical to understanding their function. Currently, functional and mechanistic characterization of lncRNAs is hampered by the limited set of research methods available[4,5]. We developed RiboPro (Ribonucleic acid proximity protein labelling) method to identify proteins associated with a target RNA *in vivo* without cross-linking or tagging RNA with aptamer sequences – both of which can lead to false positive interactions not seen under normal physiologic conditions.

Inspired by the applications of type VI Cas13 members[6,7] and proximity labeling capacity of the engineered peroxidase APEX[8], we developed the RiboPro approach, based on a catalytically dead Cas13 (dCas13) fused to APEX (Fig. 1a). In the presence of a sequence-specific guide RNA (gRNA), the dCas13-APEX fusion protein is directed to a target RNA. APEX in the chimera will biotinylate RNA proximity proteins within a few nanometers of gRNA

target sequence[9]. The biotinylated proteins can then be readily enriched using streptavidin beads and analyzed by mass spectrometry profiling or western blotting (Fig. 1a). We fused catalytically dead Cas13b from *Prevotella sp. P5-125* (dPspCas13b)[6] to the engineered soybean ascorbate peroxidase APEX2[8] with both Flag and HA tags (Fig. 1b). The expression of the fusion protein dPspCas13b-Flag-APEX2-HA (from hereon is called the RiboPro protein) was validated by western blotting using both anti-Flag and anti-HA antibodies after transfection into HEK293T cells (Fig. 1c). The RiboPro protein is expressed in both the cytoplasm and the nucleus (Fig. 1d). To verify the enzymatic activity of APEX2 in the chimera, HEK293T cells transfected with the RiboPro plasmid were treated with biotin-phenol in the presence or absence of H_2O_2. Biotinylated proteins were only detected when both biotin-phenol and H_2O_2 were added, indicating that the APEX2 retains its enzymatic activity in the RiboPro protein (Fig. 1e).

We applied RiboPro to well-known *U1* small nuclear ribonucleoprotein (snRNP) to test whether RiboPro would identify proteins known to directly bind to *U1*. Three guide RNAs (gRNAs) were designed to target different regions of *U1* small nuclear RNA (snRNA) (U1-1, U1-2, and U1-3) (Fig. 1f). Following co-transfection of *U1* gRNAs into HEK293T cells with wild-type PspCas13b, expression of *U1* snRNA was significantly downregulated compared with cells treated with a nontargeting control (NTC) gRNA, while the expression of *18S* rRNA was not affected (Fig. 1g), indicating that all three *U1* gRNAs efficiently direct PspCas13b to the target transcript. We then tested whether these *U1* gRNAs can direct the RiboPro protein to bind *U1* snRNA. In RNA immunoprecipitation experiments, the RiboPro protein was efficiently retrieved by anti-HA antibody but not isotype control IgG (Fig. 1h). *U1* snRNA was enriched around 2-fold for *U1* gRNAs compared with the NTC gRNA (Fig. 1i), indicating that the

58 RiboPro protein was selectively enriched at *U1*. U1-70k, a known *U1* direct binding partner,
59 was significantly enriched by RiboPro (Fig. 1j). *U1* proximity proteins were profiled using mass
60 spectrometry with label-free quantification and in total, 8 out of 9 *U1* direct binding proteins
61 were enriched by RiboPro (Fig. 1k and Supplementary Table 1). Together these data indicate
62 that RiboPro is able to efficiently identify direct RNA-binding proteins of a target RNA of
63 interest.

64 We next employed RiboPro to examine the RNA proximity proteome of poly (adenosine)
65 (poly[A]) tails (Supplementary Fig.1a). Since oligomers of thirty nucleotides of poly (uridine)
66 (poly[U]) are rare in the human transcriptome[10], poly(U) gRNA was used as control for poly(A)
67 gRNA. A poly(A) gRNA or poly(U) gRNA was co-transfected into HEK293T cells with the
68 RiboPro plasmid and then RiboPro was performed. Mass spectrometry profiling of two
69 independent biological experiments uncovered 910 and 1295 proteins in the poly(A) tail RNA
70 proximity proteome, and 676 and 1282 proteins in the poly(U) RNA proximity proteome,
71 respectively (Fig. 2a). In comparison to the poly(U) pulldowns, 762 and 925 proteins were
72 enriched in the two poly(A) pulldowns (Fig. 2a), 304 of which are shared by both replicates
73 (label-free quantification fold change ≥ 2 in poly[A] compared with poly[U]) (Supplementary Fig.
74 1b). In contrast, only 106 and 56 proteins were enriched in poly(U) pulldowns relative to
75 poly(A) pulldowns (Fig. 2a) with no proteins shared by both replicates when applying the same
76 threshold (Supplementary Fig. 1b). When considering subunits from the same protein complex,
77 there are 445 protein from replicate 1 and 452 proteins from replicate 2 were shared (Fig. 2b),
78 indicating the two replicates are highly reproducible at the complex level (P=1, Binomial test).
79 In contrast, in the poly(U) pulldowns, only 8 proteins from each replicate were shared (Fig. 2b),
80 suggesting poor reproducibility between replicates in our negative control (P= 2.02×10^{-05},

81 Binomial test). Only 38 proteins overlapped between the poly(A) and poly(U) RNA proximity
82 proteome, indicating these sets of proteins are distinct (P= 1.90×10^{-17}, Binomial test) (Fig. 2c).
83 Gene ontology (GO) analysis was performed on the 1347 proteins uniquely associated with
84 poly(A) tails (Supplementary Table 2). Thirteen of the top 20 enriched terms were RNA-related
85 and the majority were poly(A)+ RNA-related (Fig. 2d). Indeed, 360 proteins in poly(A) tail RNA
86 proximity proteins are poly(A)+ RNA binding proteins previously discovered in HEK293 cells[11]
87 (Fig. 2e) and 150 more poly(A)+ RNA binding proteins in HeLa or MCF7 cells[12,13]
88 (Supplementary Fig. 1c). We curated all known RBPs from literature (Supplementary Table 3,
89 see Methods and references therein) and found that 1230/1347 (91.31%) of the poly(A) tail
90 RNA proximity proteins identified by RiboPro are known RBPs (Fig. 2f) and 540/1230 (43.90%)
91 have previously been described as non-poly(A)-RBPs[14] (Supplementary Fig. 1d), suggesting
92 that more RBPs can associate with poly(A)+ RNA than previously appreciated. GO analysis of
93 117 novel poly(A) tail RNA binding proteins revealed by RiboPro suggest poly(A)+ RNA may
94 play a role in mitochondrial cristae formation, the Mediator complex and other novel processes
95 (Supplementary Fig. 1e).
96 We next examined whether known categories of poly(A)-associated proteins are identified
97 among poly(A) tail RNA proximity proteome. Polyadenylation occurs co-transcriptionally and
98 requires cleavage and polyadenylation complex (CPA) that includes cleavage and
99 polyadenylation specificity factors (CPSFs), cleavage stimulating factors (CstFs), cleavage
100 factor I and II (CFI & CFII) and poly(A) polymerase[15]. CPA members for poly(A)+RNA including
101 CPSF1, CPSF3, CstF1 and CPSF7 are present in poly(A) tail RNA proximity proteome, while
102 factors specific for poly(A)- RNA such as SLBP and ZFP100 (ZNF473) are absent (Fig. 2g).
103 Among the 12 expressed known poly(A) binding proteins, RiboPro identifies 7 (58%) (Fig. 2g).

104 The poly(A) tail is part of the 3' untranslated region (UTR), and RiboPro identifies 30 out of 41
105 (73%) expressed known 3'UTR binding proteins (Fig. 2g). In addition, poly(A) RiboPro detects
106 3 out of 8 expressed known cap binding proteins (Fig. 2g), 15 out of 19 expressed known
107 5'UTR binding proteins (Fig. 2g), and 25 translation initiation factors (Fig. 2g), including EIF4G.
108 Finally, poly(A) tail RiboPro unearths 31 out of 33 small ribosomal subunits (RPSs) and 46 out
109 of 49 large ribosomal subunits (RPLs) (Fig. 2g). These data support that 5' and 3' of poly(A)+
110 mRNA form a closed-loop structure during translation[16] and provide novel evidence for the
111 importance of poly(A) tail in mRNA translation[17]. Interestingly, the poly(A) tail RNA proximity
112 proteome also enriches 13 tRNA ligases (Supplementary Table 2), suggesting poly(A) tails
113 may have a role in tRNA charging or tRNA ligases may have unknown function in regulating
114 poly(A).
115 Adding poly(A) tails by adenylation stabilizes RNA molecules and protects transcripts from
116 degradation[17], while deadenylation triggers decapping and decay[18]. Since the poly(A) gRNA is
117 30 nt in length, poly(A) RiboPro will identify proteins proximal to any poly(A) tail over 30
118 nucleotides in length (Supplementary Fig. 1a), which will include stable transcripts, transcripts
119 actively undergoing adenylation, and transcripts undergoing deadenylation. In addition to
120 proteins related to translation (Fig. 2g), poly(A) RiboPro uncovers dozens of proteins linked
121 with RNA decay, including deadenylase complex members[19], exosome components[20],
122 nonsense-mediated decay proteins[21], decapping factors, and proteins involved in AU-rich
123 element-containing mRNA degradation[22] (Fig. 2h). The decapping enzyme and 5'-3'
124 exonuclease are not enriched in the poly(A) tail RNA proximity proteome, suggesting that
125 mRNAs no longer form closed-loop structures once the decapping process is triggered by
126 deadenylation.

127 Poly(A) tails are also important for RNA nuclear export[23], and poly(A) RiboPro detects 10
128 proteins involved in RNA nuclear export (Fig. 2i), including XPO1 and NXF1. The observation
129 that two 5' end-associated nuclear components, NCBP1 and TREX complex[24], also associate
130 with poly(A) (Fig. 2i), suggests that poly(A)+ RNA may also form closed-loop structures during
131 nuclear export. Localization signals/zip code binding protein IGF2BP1 and 9 motor proteins
132 used in mRNA transport are also present in poly(A) tail RNA proximity proteome (Fig. 2i),
133 suggesting direct involvement of the poly(A) tail in RNA transport. Furthermore, poly(A)
134 associated proteins includes microfilaments (e.g. ACTG1 and ACTN4) and membrane proteins
135 (e.g. SLC3A2 and MPZL1) related to mRNA anchorage (Fig. 2i, supplementary Table 2),
136 suggesting that the poly(A) tail participates in multiple mechanisms for RNA localization[25,26].
137 Together, these data authenticate that poly(A) RiboPro recovers *bona fide* proteins associated
138 with poly(A) tails and also reveals putative novel poly(A) RBPs not detected by alternative
139 methods.
140 Data from both *U1* and poly(A) tail RNA proximity proteomic analysis using RiboPro
141 demonstrate that this method represents a valuable new tool for probing RNA binding proteins
142 *in vivo* that is expected to shed light on the function and mechanistic roles of lncRNA in diverse
143 biological systems.
144

145 **Acknowledgements**

146 This work was supported by the National Cancer Institute (K99/R00-CA184415 and R01-
147 CA207456 to K.L.), the Cedars-Sinai Samuel Oschin Comprehensive Cancer Institute (Cancer
148 Biology Grant 231433 to K.L.) and the Ovarian Cancer Research Alliance (Ann and Sol
149 Schreiber Mentored Investigator Award 458799 to X.L.). We thank Drs. Wei Yang and Bo Zhou

150 at the Cedars-Sinai Medical Center Biomarker Discovery Platform Core for label-free
151 quantitative mass spectrometry analysis. We also thank members of the Lawrenson lab for
152 stimulating discussions.
153

154 **Competing interests**

155 The authors declare no competing interests.

156

Methods

Plasmids and cloning. pC0046-EF1a-PspCas13b-NES-HIV was a gift from Dr. Feng Zhang (Addgene plasmid # 103862). pCMV-dPspCas13b-Flag-APEX2-HA was constructed by replacing ADAR2DD-delta-984-1090 in pC0053-CMV-dPspCas13b-GS-ADAR2DD (E488Q)-delta-984-1090 (a gift from Feng Zhang, Addgene plasmid # 103869) with Flag-APEX2-HA subcloned from pcDNA3-APEX2-NES (a gift from Dr. Alice Ting, Addgene plasmid # 49386) using the following primers: dPspCas13b_F: 5'TACCCATACGATGTTCCAGATTACGCTTAAGCGGCCGCTCGAGTC3', dPspCas13b_R: 5'GTCGTCATCCTTGTAGTCGGATCCCAGTGTCAGTCTTTCAAG3', FLAG_APEX2_HA_F: 5'GACTACAAGGATGACGACG3', FLAG_APEX2_HA_R: 5'TGGAACATCGTATGGGTACTGCAGGGCATCAGCAAAC3'.

PCR was performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs, catalogue number M0491L). PCR fragments were assembled using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, catalogue number E2621S) according to manufacturer's instructions. DNA fragment was amplified from pC0043-PspCas13b crRNA backbone (a gift from Dr. Feng Zhang, Addgene plasmid # 103854) using Cas13b_crRNA-Forward: ggtgtttcgtcctttccaca and Cas13b_crRNA-Reverse: gttgtggaaggtccagtttt. Then the fragment was assembled with the following oligos after annealing using NEBuilder HiFi DNA Assembly Master Mix. Nontargeting control guide-Forward: tgtggaaaggacgaaacaccgTTTTACAACGTCGTGACTGGGAAAACCCTGgttgtggaaggtccagtttt, Nontargeting control guide-Reverse: aaaactggaccttccacaacCAGGGTTTTCCCAGTCACGACGTTGTAAAAcggtgtttcgtcctttccaca. U1 (1-30) guide-Forward: tgtggaaaggacgaaacaccgATCATGGTATCTCCCCTGCCAGGTAAGTATgttgtggaaggtccagtttt, U1 (1-30) guide-Reverse: aaaactggaccttccacaacATACTTACCTGGCAGGGGAGATACCATGATcggtgtttcgtcctttccaca. U1 (101-130) guide-Forward: tgtggaaaggacgaaacaccgCAAATTATGCAGTCGAGTTTCCCACATTTGgttgtggaaggtccagtttt, U1 (101-130) guide-Reverse: aaaactggaccttccacaacCAAATGTGGGAAACTCGACTGCATAATTTGcggtgtttcgtcctttccaca. U1(108-137) guide-Forward: tgtggaaaggacgaaacaccgACTACCACAAATTATGCAGTCGAGTTTCCCgttgtggaaggtccagtttt, U1(108-137) guide-Reverse: aaaactggaccttccacaacGGGAAACTCGACTGCATAATTTGTGGTAGTcggtgtttcgtcctttccaca. Poly (A) guide-Forward: tgtggaaaggacgaaacaccgTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTgttgtggaaggtccagtttt, Poly (A)

185 guide-Reverse: aaaactggaccttccacaacAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAcggtgtttcgtcctttccaca. Poly
186 (U) guide-Forward: tgtggaaaggacgaaacaccgAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAgttgtggaaggtccagtttt.
187 Poly (U) guide-Reverse: aaaactggaccttccacaacTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTcggtgtttcgtcctttccaca.
188 The sequences of all constructs have been confirmed using Sanger sequencing.
189
190 **Transfection and *in vivo* proximity biotinylation.** HEK293T cells were seeded into 150 mm plates and were
191 transfected with 25 µg pCMV-dPspCas13b-Flag-APEX2-HA and 15 µg Cas13b gRNAs [NTC, U1-1, U1-2, U1-3,
192 poly (A), poly (U)] using Lipofectamine 3000 (Thermo Fisher Scientific, catalog number L3000015) while ~80%
193 confluency. HEK293T were incubated with 25 mL of DMEM media containing 25µL of 500 mM biotin-phenol (Iris
194 Biotech, catalog number LS-3500.1000) in DMSO (Sigma) for 30 min at 37 °C 24h after transfection. Cells were
195 then treated with 1 mM hydrogen peroxide (H_2O_2) (Sigma-Aldrich, catalog number H1009) for 1 min on a
196 horizontal shaker at room temperature. The labeling solution was aspirated, and cells were washed twice with 25
197 mL of quencher solution (10 mM sodium azide [Sigma-Aldrich, catalog number S2002-5G], 10 mM sodium
198 ascorbate [Sigma-Aldrich, catalog number PHR1279-1G], and 5 mM Trolox [Sigma-Aldrich, catalog number
199 238813-1G] in Dulbecco's phosphate buffed saline [DPBS, Thermo Fisher Scientific, catalog number 14040182]).
200 Cells were then washed three times with 15 mL of DPBS and were pelleted by centrifugation at 1,500×g for 5 min
201 at 4 °C. Cell pellets were snap frozen and stored at −80 °C.
202
203 **Streptavidin enrichment of biotinylated proteins.** Cell pellets from two 150 mm plates of transfected HEK293T
204 cells were lysed in 2 mL cell lysis buffer (10 mM HEPES, pH7.5 by KOH, 150 mM NaCl, 0.1% NP-40, 5 mM
205 EGTA, 5 mM Trolox, 10 mM Sodium ascorbate acid, 10 mM Sodium azide, 1 mM PMSF). Streptavidin magnetic
206 beads (Thermo Fisher Scientific, catalog number 88817) were washed twice with cell lysis buffer and 3.5 mg of
207 each whole cell lysate sample were incubated with 100 µL magnetic bead slurry with rotation for 2 h at room
208 temperature. After enrichment, the flowthrough was removed and beads were washed with 2×1 mL cell lysis
209 buffer, 1mL 1 M KCl, 1 mL 0.1 M Na_2CO_3, 1 mL of 2 M urea in 10 mM Tris-HCl (pH 8.0), and again with 2×1 mL
210 cell lysis buffer. These denaturing washes are important to disrupt protein-protein interactions and ensure
211 enrichment only of proteins directly biotinylated by APEX2. Biotinylated proteins were then eluted by boiling the

212 magnetic beads in 30 μL 4×Laemmli sample buffer (Bio-Rad, catalog number 1610747) supplemented with 20
213 mM DTT and 2 mM biotin.
214
215 **Label-free quantitative mass spectrometry analysis**
216 The proteins enriched by RiboPro were analyzed by label-free quantitative mass spectrometry analysis as
217 previously described[27] at Cedars-Sinai Medical Center Biomarker Discovery Platform Core.
218
219 **Comparison of RBP databases**
220 The proteins annotated as RBPs used in our study were retrieved from corresponding manuscripts [11–14,28–44]. A list
221 of human RBPs (GO_RNA Binding) were retrieved from QuickGO (https://www.ebi.ac.uk/QuickGO/) via searching
222 "RNA Binding" (GO:0003723) and select "*Homo sapiens*" under Taxon. Other categories of human RBPs were
223 retrieved in the same way by searching corresponding terms. The venn diagrams were generated using online
224 tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) and pseudovenn diagrams were generated using R
225 package UpSetR.
226
227 **GO analysis**
228 GO analyses were performed using Metascape [45](http://metascape.org/gp/index.html#/main/step1).
229
230 **Cellular fractionation.** Six million HEK293T cells were treated with a plasma membrane lysis buffer (10 mM Tris-
231 HCl, pH 7.5, 0.15% NP-40, 150 mM NaCl) on ice for 4 min after homogenization by flicking. Lysates were loaded
232 onto a 24% sucrose cushion (24% RNase-free sucrose in plasma membrane lysis buffer) using large orifice tips,
233 and centrifuged at 15,000 × *g* for 10 min at 4°C. The supernatant (cytoplasmic fraction) was retained after
234 centrifugation, and the pellet (nuclear fraction) was washed with 1×PBS/1 mM EDTA and resuspended in 200 μL
235 of 1×PBS/1 mM EDTA. Fractionation efficiency was validated by Western blot using β-tubulin (Sigma Aldrich,
236 catalogue number T8328, 1:2,000) as a cytoplasmic marker and U1-70k (EMD Millipore, catalogue number 05-
237 1588, 1:1,000) as a nuclear marker.
238

239 **RNA Immunoprecipitation (RIP).** Twelve microliter Dynabeads Protein A (Thermo Fisher Scientific, catalog
240 number: 10001D) were washed with 200 μL HBS (150 mM NaCl, 10 mM HEPES, pH7.5 by KOH and incubated
241 with 2 μg anti-HA (Santa Cruz, catalogue number sc-7392) or 2 μg rabbit IgG isotype control (Thermo Fisher
242 Scientific, catalogue number 10500C) in the presence of 80 μL HBS buffer at room temperature for 1 h. Eight
243 million HEK293T cells were lysed with 800 μL cell lysis buffer (10 mM HEPES, pH7.5 by KOH, 150 mM NaCl,
244 0.1% NP-40, 5 mM EGTA, supplemented with 1× protease inhibitor cocktail and 1mM PMSF) at 4°C for 1 h. Cell
245 debris and insoluble proteins were removed by centrifugation at 4°C, 12,000 × *g* for 10 min, and the supernatants
246 were incubated with HA-conjugated Dynabeads at 4°C for 1 h. The Dynabeads were then washed 3 times with
247 wash buffer (150 mM NaCl, 10 mM HEPES, pH7.5 by KOH, 0.1% NP-40). Proteins were eluted with 22 μL
248 1×Laemmli sample buffer (Bio-Rad, catalog number: 1610747) by boiling for western blotting.
249
250 **RNA extraction and RT-qPCR.** Immunoprecipitated RNA associated with dPspCas13b-Flag-APEX2-HA was
251 extracted using TRIzol LS (Thermo Fisher Scientific, catalogue number: 10296028). M-MLV reverse transcriptase
252 (Promega, catalogue number: M5301) and random hexamers (Promega, catalogue number: C1181) were used
253 for reverse transcription. Gene expression was quantified by RT-qPCR using iQ SYBR Green supermix (Bio-Rad,
254 catalogue number: 170-8886). The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method and
255 normalized to GAPDH. Five nanograms cDNA was used for RT-qPCR analysis on CFX96 Touch Real-Time PCR
256 Detection System (Bio-Rad) using the following primer pairs: U1 snRNA-RT-Forward:
257 5'CCAGGGCGAGGCTTATCCATT3', U1 snRNA-RT-Reverse: 5'GCAGTCCCCCACTACCACAAAT3', 18S rRNA-
258 RT-Forward: 5'CAGCCACCCGAGATTGAGCA3', 18S rRNA-RT-Reverse: 5'TAGTAGCGACGGGCGTGTG3',
259 GAPDH-RT-Forward: 5'TGCCAAATATGATGACATCAAGAA3', GAPDH-RT-Reverse:
260 5'GGAGTGGGTGTCGCTGTTG3'.
261
262 **Western blotting.** Protein samples were run on 4-20% gradient precast protein gel (Bio-Rad, catalogue number:
263 456-1096) and then transferred onto PVDF membrane (Bio-Rad, catalogue number: 1704157). After one hour
264 blocking, membranes were incubated with anti-Flag (Santa Cruz, catalogue number sc-166384), anti-HA (Santa
265 Cruz, catalogue number sc-7392), anti-biotin (Santa Cruz, catalogue number sc-57636), or anti- β-actin (Santa
266 Cruz, catalogue number sc-47778, 1:2,000) at 4°C overnight. Membranes were washed three times with Tris-

267 buffered saline containing 0.5% Tween 20 (TBST) before incubating with HRP-conjugated secondary antibody at
268 room temperature for 2 h. Then the membranes were incubated briefly with ECL Western Blotting substrate
269 (Thermo Fisher Scientific, catalogue number: 32106) after three times wash with TBST. The membranes were
270 exposed to HyBlot Autoradiography Film (Denville Scientific, catalogue number: E3018).
271

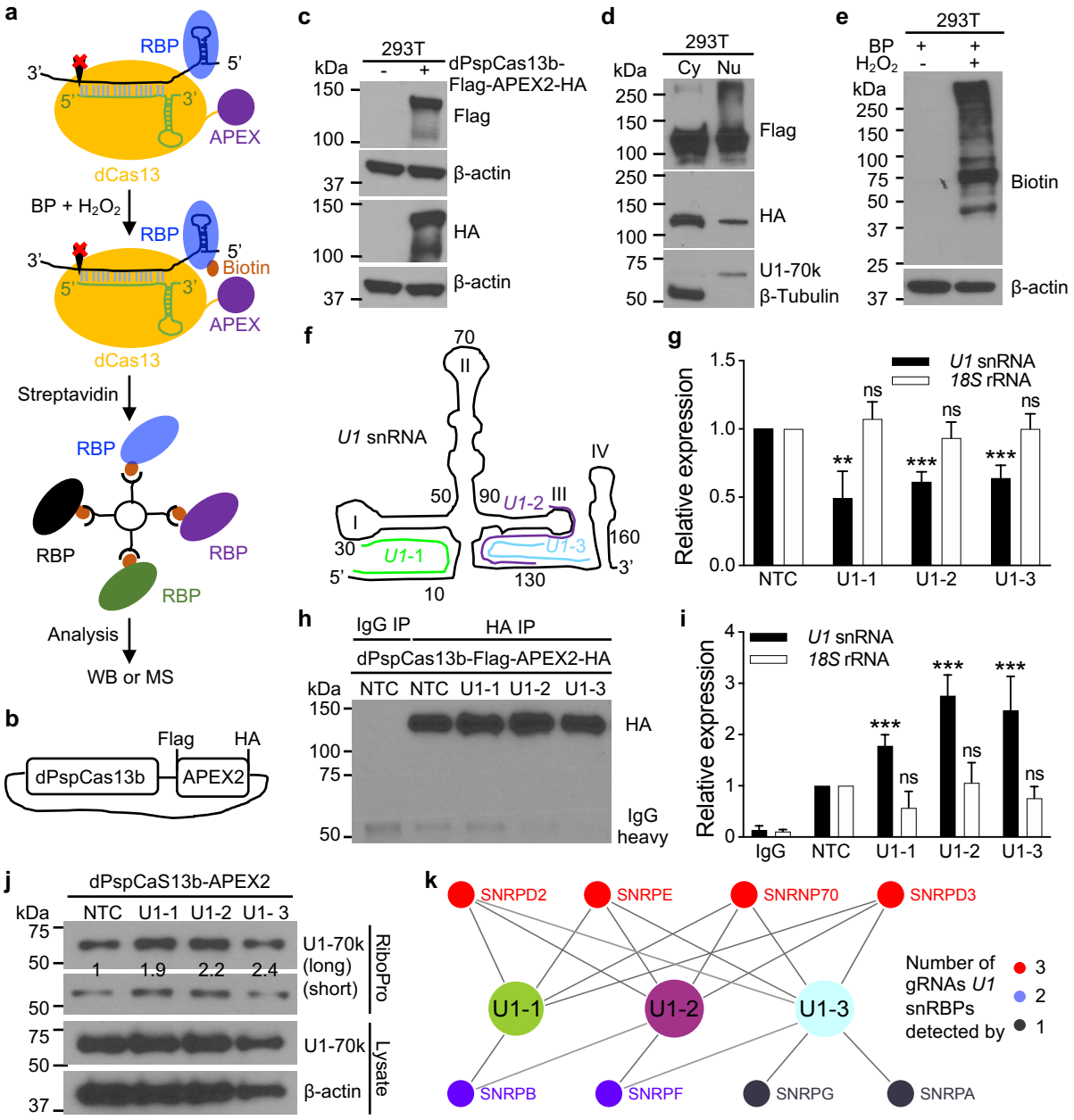
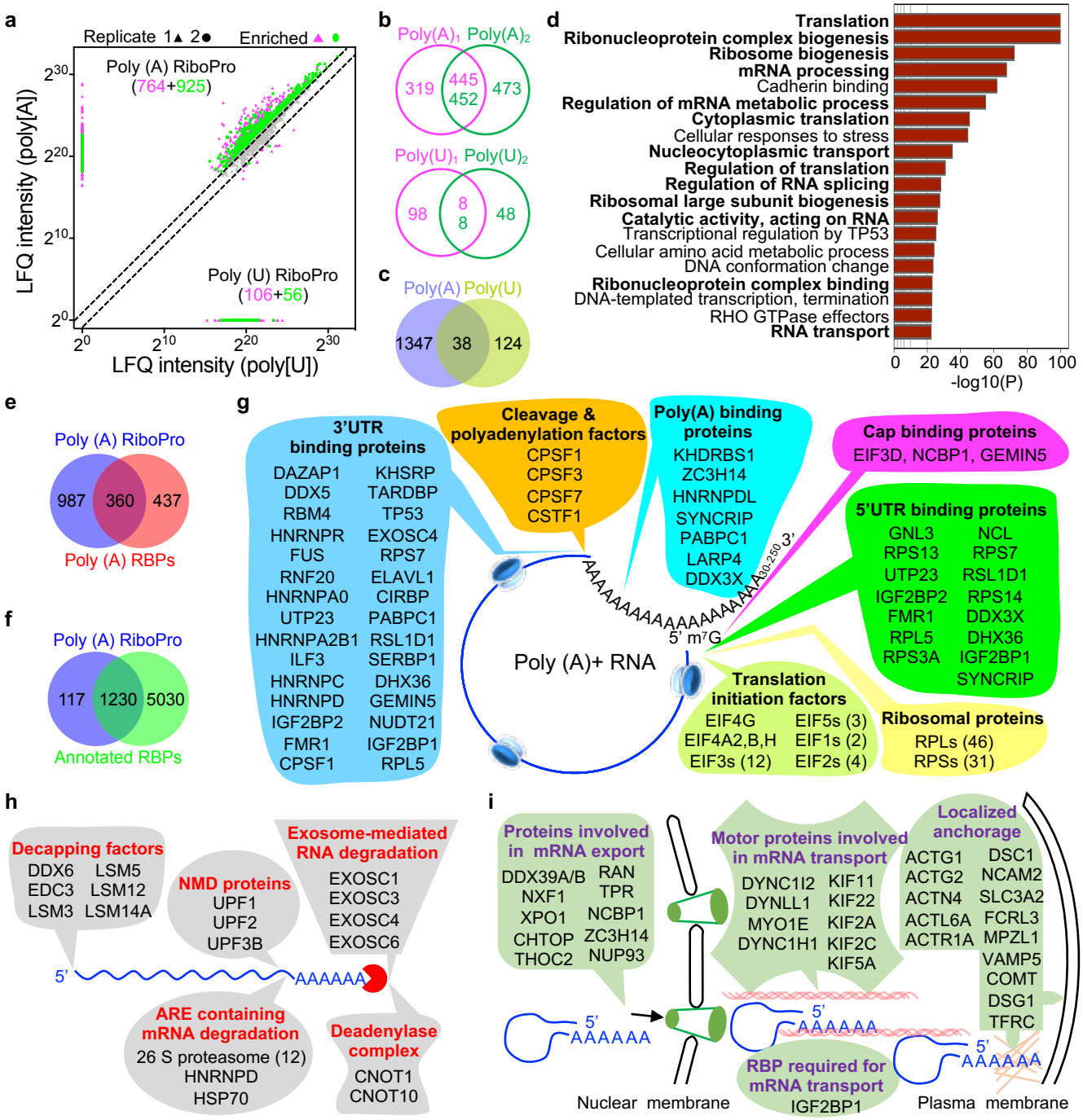


Figure Legend

Figure 1. Proof of principle for RiboPro method

a. Schematic representation of RiboPro approach. A sequence-specific gRNA directs dPspCas13-APEX to the target RNA and APEX in the chimera biotinylates proximal proteins in the presence of biotin-phenol (BP) and H_2O_2. Biotinylated proteins are isolated by streptavidin dynabeads and subsequently analyzed by mass spectrometry (MS) or western blotting

279 (WB). **b**. Diagram of dPspCas13b-Flag-APEX2-HA construct. **c**. Expression validation of
280 dPspCas13b-Flag-APEX2-HA by WB. **d**. dPspCas13b-Flag-APEX2-HA is expressed in both
281 cytoplasm and nucleus. **e**. Functional validation of dPspCas13b-Flag-APEX2-HA. HEK293T
282 cells transfected with dPspCas13b-Flag-APEX2-HA were treated with BP or BP+H_2O_2. Whole
283 cell lysates were blotted with anti-biotin antibody. β-actin was blotted as a loading control. **f**.
284 Designing of 3 gRNAs targeting *U1* snRNA. **g**. Verification of *U1* snRNA gRNAs. HEK293T
285 cells were transfected with PspCas13b and NTC gRNA or *U1* snRNA gRNAs (1:1 ratio). The
286 expression of *U1* snRNA or nontargeting *18S* rRNA was quantified by RT-qPCR. **h**. Western
287 blotting confirms successful pulldown of dPspCas13b-Flag-APEX2-HA by RIP. HEK293T cells
288 were transfected with dPspCas13b-Flag-APEX2-HA and NTC gRNA or *U1* snRNA gRNAs
289 (equal ratio). HA antibody or isotype control IgG were used to immunoprecipitate dPspCas13b-
290 Flag-APEX2-HA and total RNA was extracted and measured by RT-qPCR. **i**. RT-qPCR
291 measures the abundance of *U1* snRNA and *18S* rRNA from RIP. **j**. U1-70k was enriched by
292 specific gRNAs for *U1* snRNA. HEK293T cells transfected with pCMV-dPspCas13b-Flag-
293 APEX2-HA and NTC gRNA or *U1* snRNA gRNAs were treated with BP+H_2O_2. Whole cell
294 lysates (Input) or streptavidin enriched biotinylated proteins (RiboPro) were blotted. Short and
295 long exposures are shown for U1-70k. **k**. MS profiling reveals additional *U1* snRNA direct
296 binding proteins enriched by RiboPro.
297



298

299 **Figure 2. RiboPro reveals human poly (A)+ tail RNA proximity proteome**

300 **a.** Correlation of label-free quantification (LFQ) of proteins identified by RiboPro of poly (A) or

301 poly (U) tail from two independent replicates. R^2=0.24, P=2.12×10^{-5}, *Pearson*'s correlation

302 using shared candidates for both replicates. **b.** Shared proteins identified by two replicates. **c.**

303 Intersecting proteins between poly (A) and poly (U) protein proximitome. $P = 1.90 \times 10^{-17}$,

304 Binomial test. **d.** GO analysis of poly (A) tail RNA proximity proteome. Top 20 most

305 significantly enriched GO terms are shown. **e.** Comparison of poly (A) RiboPro with poly (A)

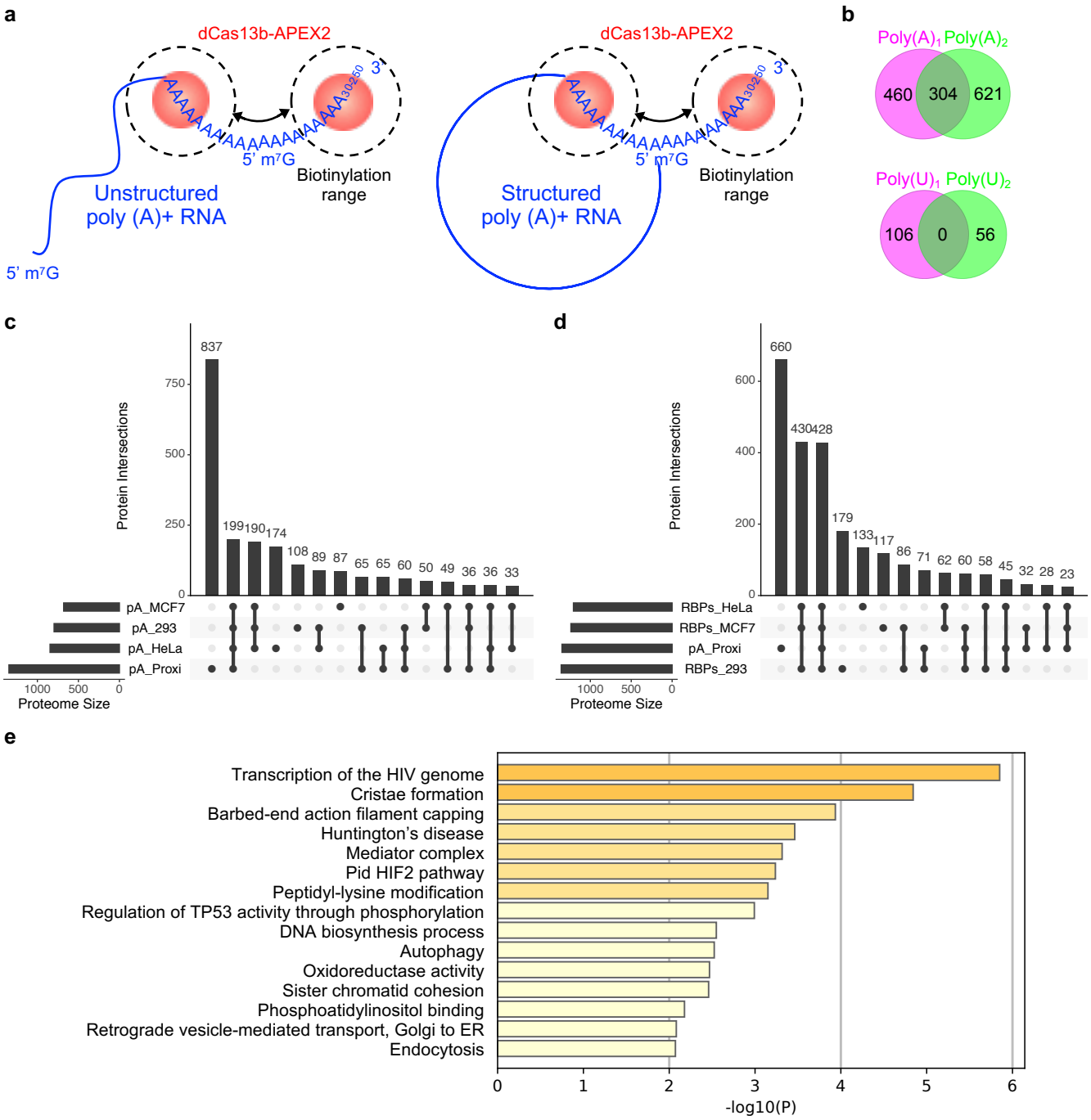
306 RNA binding proteins in HEK293 cells. **f.** Comparison of poly (A) RiboPro with all annotated

307 RBPs. **g.** Poly (A) RiboPro reveals known poly (A) RNA binding proteins for poly (A)+ mRNAs.

308 **h.** Poly (A) RiboPro uncovers proteins related to RNA turnover. **i.** Poly (A) RiboPro identifies

309 proteins related to RNA nuclear export, RNA transport, and RNA localization.

310



Supplementary Figure 1. Poly(A) tail RNA proximity proteome revealed by RiboPro

a. RiboPro of poly(A) tail RNA proximity proteome. In the presence of guide RNA, RiboPro protein is directed to poly (A) tail ranging from 30 nt to 250 nt. Two hypothetical scenarios are shown for poly (A)+ RNA, under which RiboPro should identify different categories of RNA

316 proximity proteins. In the unstructured scenario, RiboPro is likely to detect poly (A) RNA
317 binding proteins and 3'UTR binding proteins that bind proximal to poly (A) tail as the RNA
318 proximity proteome of poly (A) tail. In addition to these two categories of proteins, RiboPro will
319 likely identify cap-binding proteins, 5'UTR binding proteins that bind proximal to the cap in the
320 structured scenario. **b.** Intersected proteins between two biological replicates for both poly (A)
321 and poly (U) tail RNA proximity proteome. **c.** Pseudovenn diagram using UpSetR
322 demonstrates the intersections among poly (A) tail RNA proximity proteome, poly (A) RNA
323 binding proteomes in HEK293 cells, HeLa cells, and MCF7 cells, respectively. **d.** Pseudo-venn
324 diagram demonstrates the intersections among poly (A) tail RNA proximity proteome, total
325 RNA binding proteomes including both poly (A) RNA and non-poly (A) RNA binding proteins in
326 HEK293 cells, HeLa cells, and MCF7 cells, respectively. **e.** Enriched GO terms for unique poly
327 (A) tail RNA proximity proteins.
328

329 **References**

330 1. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914
331 (2011).
332 2. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–
333 46 (2012).
334 3. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional
335 gene regulation. *FEBS Lett.* **582**, 1977–1986 (2008).
336 4. Chu, C., Spitale, R. C. & Chang, H. Y. Technologies to probe functions and mechanisms of long
337 noncoding RNAs. *Nat. Struct. Mol. Biol.* **22**, 29–35 (2015).
338 5. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein interactions. *Nat.*
339 *Methods* **16**, 225–234 (2019).
340 6. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
341 7. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017).
342 8. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat.*
343 *Methods* **12**, 51–54 (2015).
344 9. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* (2019)
345 doi:10.1016/j.cell.2019.05.027.
346 10. Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length
347 and 3′ End Modifications. *Molecular Cell* **53**, 1044–1052 (2014).
348 11. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding
349 Transcripts. *Molecular Cell* **46**, 674–690 (2012).

350 12. Milek, M. *et al.* DDX54 regulates transcriptome dynamics during DNA damage response. *Genome*
351 *Res.* **27**, 1344–1359 (2017).
352 13. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.
353 *Cell* **149**, 1393–1406 (2012).
354 14. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest.
355 *Cell* **176**, 391-403.e19 (2019).
356 15. Shi, Y. & Manley, J. L. The end of the message: multiple protein–RNA interactions define the mRNA
357 polyadenylation site. *Genes Dev.* **29**, 889–897 (2015).
358 16. Vicens, Q., Kieft, J. S. & Rissland, O. S. Revisiting the Closed-Loop Model and the Nature of mRNA 5′
359 –3′ Communication. *Molecular Cell* **72**, 805–812 (2018).
360 17. Dreyfus, M. & Régnier, P. The Poly(A) Tail of mRNAs. *Cell* **111**, 611–613 (2002).
361 18. Norbury, C. J. Cytoplasmic RNA: a case of the tail wagging the dog. *Nat. Rev. Mol. Cell Biol.* **14**, 643–
362 653 (2013).
363 19. Collart, M. A. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley*
364 *Interdiscip Rev RNA* **7**, 438–454 (2016).
365 20. Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: The exosome.
366 *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**, 552–560 (2013).
367 21. Chang, Y.-F., Imam, J. S. & Wilkinson, M. F. The Nonsense-Mediated Decay RNA Surveillance
368 Pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007).
369 22. Laroia, G., Cuesta, R., Brewer, G. & Schneider, R. J. Control of mRNA decay by heat shock-ubiquitin-
370 proteasome pathway. *Science* **284**, 499–502 (1999).

371 23. Huang, Y. & Carmichael, G. G. Role of polyadenylation in nucleocytoplasmic transport of mRNA.
372 *Mol. Cell. Biol.* **16**, 1534–1542 (1996).
373 24. Cheng, H. *et al.* Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389–
374 1400 (2006).
375 25. Mofatteh, M. & Bullock, S. L. SnapShot: Subcellular mRNA Localization. *Cell* **169**, 178-178.e1 (2017).
376 26. Holt, C. E. & Bullock, S. L. Subcellular mRNA Localization in Animal Cells and Why It Matters. *Science*
377 **326**, 1212–1216 (2009).
378 27. Zhou, B. *et al.* Low-Background Acyl-Biotinyl Exchange Largely Eliminates the Coisolation of Non- *S* -
379 Acylated Proteins and Enables Deep *S* -Acylproteomic Analysis. *Anal. Chem.* **91**, 9858–9866 (2019).
380 28. Bao, X. *et al.* Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15**, 213–220
381 (2018).
382 29. Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved
383 enigmRBPs. *Nat Commun* **6**, 10127 (2015).
384 30. Brannan, K. W. *et al.* SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-
385 Protein Interactomes. *Molecular Cell* **64**, 282–293 (2016).
386 31. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular*
387 *Cell* **63**, 696–710 (2016).
388 32. Caudron-Herger, M. *et al.* R-DeeP: Proteome-wide and Quantitative Identification of RNA-
389 Dependent Proteins by Density Gradient Ultracentrifugation. *Mol. Cell* **75**, 184-199.e10 (2019).
390 33. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding
391 specificities. *Nucleic Acids Research* **39**, D301–D308 (2011).

392 34. Conrad, T. *et al.* Serial interactome capture of the human cell nucleus. *Nat Commun* **7**, 11212
393 (2016).
394 35. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.*
395 **15**, 829–845 (2014).
396 36. He, C. *et al.* High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of
397 Embryonic Stem Cells. *Molecular Cell* **64**, 416–430 (2016).
398 37. Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol.*
399 *Biol.* **20**, 1122–1130 (2013).
400 38. Mullari, M., Lyon, D., Jensen, L. J. & Nielsen, M. L. Specifying RNA-Binding Regions in Proteins by
401 Peptide Cross-Linking and Affinity Purification. *J. Proteome Res.* **16**, 2762–2772 (2017).
402 39. Perez-Perri, J. I. *et al.* Discovery of RNA-binding proteins and characterization of their dynamic
403 responses by enhanced RNA interactome capture. *Nat Commun* **9**, 4408 (2018).
404 40. Queiroz, R. M. L. *et al.* Comprehensive identification of RNA-protein interactions in any organism
405 using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* **37**, 169–178 (2019).
406 41. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–
407 177 (2013).
408 42. Sundararaman, B. *et al.* Resources for the Comprehensive Discovery of Functional RNA Elements.
409 *Molecular Cell* **61**, 903–913 (2016).
410 43. Treiber, T. *et al.* A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis.
411 *Molecular Cell* **66**, 270-284.e13 (2017).
412 44. Urdaneta, E. C. *et al.* Purification of cross-linked RNA-protein complexes by phenol-toluol
413 extraction. *Nat Commun* **10**, 990 (2019).

414 45. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level
415 datasets. *Nat Commun* **10**, 1523 (2019).
416