Deciphering human ribonucleoprotein regulatory networks

Neelanjan Mukherjee[1], Hans-Hermann Wessels[2], Svetlana Lebedeva[2], Marcin Sajek[3,4], Mahsa Ghanbari[2], Aitor Garzia[4], Alina Munteanu[2], Thalia Farazi[4*], Jessica I Hoell[4,5], Kemal Akat[4], Thomas Tuschl[4], Uwe Ohler[2]

[1]Department of Biochemistry and Molecular Genetics, RNA Bioscience Initiative, University of Colorado School of Medicine, Aurora, Colorado 80045, USA.

[2]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany.

[3]Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland

[4]Howard Hughes Medical Institute and Laboratory for RNA Molecular Biology, The Rockefeller University, 1230 York Ave, Box 186, New York, NY 10065

[5]Department of Pediatric Oncology, Hematology and Clinical Immunology, Center for Child and Adolescent Health, Medical Faculty, Heinrich Heine University of Dusseldorf, Dusseldorf, Germany

[*]Current address: Juno Therapeutics, 400 Dexter Avenue North, Seattle, WA 98109

Correspondence to:

ttuschl@rockefeller.edu, uwe.ohler@mdc-berlin.de, neelanjan.mukherjee@ucdenver.edu

RNA-binding proteins (RBPs) control and coordinate each stage in the life cycle of RNAs. Although *in vivo* binding sites of RBPs can now be determined genome-wide, most studies typically focused on individual RBPs. Here, we examined a large compendium of 114 high-quality transcriptome-wide *in vivo* RBP-RNA cross-linking interaction datasets generated by the same protocol in the same cell line and representing 64 distinct RBPs. Comparative analysis of categories of target RNA binding preference, sequence preference, and transcript region specificity was performed, and identified potential posttranscriptional regulatory modules, i.e. specific combinations of RBPs that bind to specific sets of RNAs and targeted regions. These regulatory modules encoded functionally related proteins and exhibited distinct differences in RNA metabolism, expression variance, as well as subcellular localization. This integrative investigation of experimental RBP-RNA interaction evidence and RBP regulatory function in a human cell line will be a valuable resource for understanding the complexity of post-transcriptional regulation.

35 **Introduction**

36 Of the 20,345 annotated protein-coding genes in human, at least 1,542 are RNA-binding proteins
37 (RBPs) (Gerstberger et al., 2014). RBPs interact with RNA regulatory elements within RNA
38 targets to control splicing, nuclear export, localization, stability, and translation (Moore, 2005).
39 RBPs have specificity to bind one or multiple RNA categories, including messenger RNA
40 (mRNA) and diverse categories of non-coding RNA such as ribosomal RNA (rRNA), transfer
41 RNA (tRNA), small nuclear and nucleolar RNA (snRNA/snoRNA), microRNA (miRNA), and
42 long non-coding RNA (lncRNA). Mutations in RBPs or RNA regulatory elements can result in
43 defects in RNA metabolism that cause human disease (Cooper et al., 2009; Fredericks et al.,
44 2015).
45
46 A standard technique for *in vivo* global identification of RBP-RNA interaction sites consists of
47 immunoprecipitating the ribonucleoprotein (RNP) complex, isolating the bound RNA, and
48 quantifying the RNA targets by microarrays or deep sequencing (Tenenbaum et al., 2000; Zhao
49 et al., 2010). The introduction of cross-linking prior to immunoprecipitation (CLIP) as well as
50 RNase digestion enabled the biochemical mapping of individual interaction sites (Ule et al.,
51 2003). Subsequent modifications to CLIP increased the resolution of the interaction sites (Hafner
52 et al., 2010; König et al., 2010). One of these methods, photoactivatable ribonucleoside-
53 enhanced cross-linking and immunoprecipitation (PAR-CLIP), utilizes 4-thiouridine or 6-
54 thioguanosine combined with 365 nm UV crosslinking to produce single-nucleotide RBP-RNA
55 interaction evidence that is utilized to define binding sites (Corcoran et al., 2011; Garzia et al.,
56 2017b; Hafner et al., 2010).

57 Experimentally-derived RBP binding sites provide valuable functional insights. First, they can
58 reveal the rules for regulatory site recognition by the RBP, whether due to sequence and/or
59 structural characteristics. Second, the region and position of the interaction sites of an RBP
60 within transcripts provides insights into its role in RNA metabolism and its subcellular
61 localization. For example, if most of the mapped interaction sites are intronic and adjacent to
62 splice sites, the RBP is highly likely to be a nuclear splicing factor rather than a cytoplasmic
63 translation factor. Finally, these data reveal the target transcripts and therefore the potential
64 biological role for the RBP.
65
66 Throughout the life of an RNA, interactions with many different RBPs determine the ultimate
67 fate of the transcript. Even though profiling of the interaction sites of a single RBP is clearly
68 powerful, it does not provide information on other RBPs potentially targeting the same RNA or
69 on other regulatory elements within the RNA. Small comparative efforts focusing on the
70 regulation of splicing, 3' end processing, RNA stability by AU-rich elements, and miRNA-
71 mediated silencing have demonstrated the value of integrating interaction sites from multiple
72 RBPs (Martin et al., 2012; Mukherjee et al., 2014; Pandit et al., 2013; Zhang et al., 2010).
73 Therefore, a large-scale comparative examination of interaction sites for many RBPs will yield
74 valuable knowledge regarding the architecture and determinants of RNA regulatory networks.
75
76 At least 173 PAR-CLIP experiments have been performed in HEK293 cells to date, laying the
77 groundwork for a large-scale integrative analysis and complementing efforts of ENCODE, which
78 focused on other cell types and utilized other CLIP protocols (Van Nostrand et al., 2016). We
79 describe a concerted effort to identify and uniformly process all high-quality PAR-CLIP data sets

80 by evaluating the characteristic T-to-C transitions induced by photocrosslinking. Using the
81 resulting compendium of high-quality in vivo RBP interaction maps from the same cell line
82 enabled us to determine the relationship between RBPs with respect to their preferred category of
83 target RNA and any underlying sequence specificity. We uncovered regulatory modules reflected
84 by combinatorial binding events, and assessed their role and functional implications on RNA
85 metabolism. Finally, our results support the role of RBPs in buffering gene expression variance.
86

87 **Results**

88 **A high-quality map of in vivo RBP-RNA interactions across 64 proteins**

89 In order to generate a comprehensive quantitative resource of RBP-RNA interactions within a
90 human cell line, we identified 166 published PAR-CLIP data sets performed predominantly in
91 HEK293 cells, and added 7 new libraries generated in our laboratories (Sup Table 1). Typically,
92 these datasets were generated using transgenic HEK293 cell lines in which each individual RBP
93 was FLAG-tagged and recombined into the same chromosomal locus containing a strong
94 promoter. In this way, the expression of each RBP as well as the strength of its
95 immunoprecipitation were generally comparable. Furthermore, the availability of orthogonal
96 transcriptome-wide datasets quantifying individual steps of RNA metabolism made HEK293
97 cells ideal for examining the functional characteristics of RNA targets (Mukherjee et al., 2017).

98

99 Each of the 173 PAR-CLIP libraries generated in HEK293 were subject to a stringent analysis
100 strategy to retain high-quality datasets (Supplemental Table 1). First, each library was analyzed
101 using the PAR-CLIP Suite v1.0 (https://rnaworld.rockefeller.edu/PARCLIP_suite) (Garzia et al.,

102 2017b) to discriminate significant target RNA categories from non-crosslinked background RNA
103 categories populated by fragments of abundant cellular RNAs (see Methods, Supplemental Fig.
104 1A). Next, we defined binding sites based on the local density of T-to-C transitions using
105 PARpipe (https://github.com/ohlerlab/PARpipe) (Corcoran et al., 2011) and only retained those
106 libraries with sufficiently high read counts and T-to-C transition specificity compared to a deeply
107 sequenced background reference library (Supplemental Fig 1b) (Friedersdorf and Keene, 2014).
108 Since the immunoprecipitation step was omitted in this reference library it served as an effective
109 comparison point to score read count and T-to-C transition for all RBPs. Finally, for RBPs with
110 more than 3 libraries available, outlier libraries exhibiting poor correlation of 6-mer frequencies
111 were excluded (Supplemental Fig 1d, e). This resulted in 114 libraries corresponding to 64 RBPs
112 that were the basis for downstream analysis. There were eight RBP families represented by two
113 or more RBPs.

114

115 **Grouping RBPs by annotation category and positional binding site preferences**

116 As first step to describe RBP-RNA regulatory networks, we determined the relative binding
117 preference of each RBP for specific target RNA annotation categories (Supplemental Table 2).
118 For each library, we calculated an RNA annotation category preference value, defined as the
119 difference in the fraction of T-to-C reads per annotation category between each RBP library and
120 the reference library. We performed hierarchical clustering of RBPs by annotation category
121 preference, using Ward's method and Euclidean distances. This yielded eight clusters of binding
122 preference (Figure 1a – orange line demarcates cluster definitions) with varying enrichment or
123 depletion for individual or combinations of specific annotation categories. For each of these
124 clusters, we compiled a detailed table summarizing the reported functions for each of the RBPs

125 (Table 1). Taken together, clustering by RNA annotation category separated RBPs into groups
126 according to their known subcellular localization and functions.

127 Three of the eight clusters (clusters 2, 4, and 5) contained nine RBPs that exhibited preference
128 for categories of non-coding RNA (rRNA, snRNA, snoRNA, and tRNA), but not mRNA,
129 precursor mRNA (pre-mRNA), or lncRNA. The remaining five clusters contained 55 RBPs
130 exhibiting preference for binding to mRNA, pre-mRNA and long-noncoding RNA (lncRNA)
131 annotation categories. The RBPs in clusters 1, 6, 7, and 8 exhibited strong preferences for
132 various mRNA annotation categories. The RBPs in cluster 3 did not exhibiting strong preference
133 for specific mRNA annotation categories. Additionally, for each of the RBPs in the cluster, we
134 performed a positional meta-analysis of binding sites with respect to major transcript landmarks
135 within target mRNAs. Many of the RBPs also showed strong preferences for binding to specific
136 positions within mRNAs relating to their role in specific steps of mRNA processing (Table 1).

137 We hypothesized that target annotation category preferences and positional binding preferences
138 should reflect subcellular localization of the RBP and its role(s) in mRNA processing. Cluster 6
139 contained twelve RBPs and exhibited strong preference for intronic regions and to a lesser
140 degree 3' UTRs of mRNAs and lncRNAs. The intronic preference was consistent with the
141 predominantly nuclear localization of these RBPs and the pre-mRNA splicing process. ELAVL1,
142 which is the sole member of the ELAVL1 family of RBPs that is predominantly localized in the
143 nucleus but capable of shuttling to the cytoplasm, exhibited positional binding flanking the end
144 of the 3' UTR and for 5' and 3' splice sites. Cluster 8 contained fourteen RBPs and exhibited
145 distinct preference for 3' UTR regions. This included the unpublished and predominantly
146 cytoplasmic ELAVL1 family members, ELAVL2, ELAVL3, and ELAVL4, which exhibited a
147 strong positional preference for binding in the distal region of the 3' UTR and acting

148 predominantly on mature mRNA (Mansfield and Keene, 2012). In summary, the annotation
149 category preferences and positional binding preferences implicated the specific steps of mRNA
150 processing the RBPs potentially regulate.

151

152 **The spectrum of RNA sequence specificity**

153 RBPs exist on a spectrum of specificity depending on a variety of primary and secondary
154 structure features (Jankowsky and Harris, 2015). Here, our goal was to identify the RBPs with
155 substantial primary sequence specificity and then examine their sequence preference. For each of
156 the 55 RBPs, we counted all possible 6-mers using Jellyfish (Marçais and Kingsford, 2011) for
157 the reads contributing to PARalyzer-defined binding sites. We observed 6-mer frequencies
158 ranging as high as 512-fold to as low as 5-fold over a uniform distribution of 6-mers
159 (Supplemental figure 2a). In contrast, our reference background library exhibited 16-fold
160 enrichment of at least one 6-mer compared to uniform. AGO1-4 libraries were excluded from 6-
161 mer analysis due to the overwhelming sequence contribution from crosslinked miRNAs. Twenty-
162 seven RBPs did not have a single 6-mer found at higher frequency than present in the reference
163 sample. Amongst these RBPs established or expected to display low sequence-specificity were
164 the RNA helicase MOV10, the nuclear exosome component DIS3, and the EIF3 complex
165 translation initiation factors.

166

167 For each of the 24 RBPs with stronger sequence enrichment than the reference library, we
168 clustered the top 5 sequences enriched over the reference library (Figure 2). Our results
169 recapitulated the sequence preference for the RBPs in this group with well-characterized
170 sequence motifs (detailed in Table 2). The ELAVL1 family proteins, which bound to different

171 regions and positions of mRNA, showed similar preference for U- and AU-rich 6-mers, while
172 ZFP36 only enriched a subset of the AU-rich 6-mers (Mukherjee et al., 2014). Complementing
173 the 6-mer enrichment analysis, we performed motif analysis for each RBP library with the motif
174 finding algorithm SSMART (sequence-structure motif identification for RNA-binding
175 proteins, (Munteanu et al., 2018)) (Supplemental Fig 2b). For most RBPs, we observed strong
176 concordance between the two analyses. RBM20 was a clear exception, for which we observed
177 the established UCUU-containing motifs (Maatz et al., 2014) with SSMART, but a GA-rich
178 sequence in the 6-mer enrichment analysis. However, we do observe UCUU-containing motifs in
179 the top 15, but not top5 6-mers for RBM20. Altogether, our analysis was remarkably consistent
180 with previously reported motifs in spite of differences in data processing and analysis (detailed
181 Table 2).

182

183 **Identification of RNA regulatory modules**

184 To understand the functional impact of co-regulation by multiple RBPs, we analyzed the co-
185 variation in binding patterns of all 55 RBPs across 13,299 target RNA encoding genes to probe
186 for the existence of regulatory modules, i.e., specific subsets of RNAs implicated in similar
187 function bound by subsets of RBPs. To this end, we employed Factor Analysis (FA), which
188 reduces a large number of observed variables to a smaller number of latent *factors*. Here, our
189 observed variables represented the normalized RBP binding (see methods) for each of the 55
190 RBPs across all target RNA encoding genes (n=13,299). The latent *factors* represented similar
191 binding patterns to RNA targets by one or more of the 55 RBPs. RBPs exhibiting high loadings
192 for the same *factor* would have very similar binding patterns to RNA targets. Importantly in this

193 framework, a single RBP could be assigned to multiple *factors*, just as a single RBP can
194 participate in multiple RNPs and regulate different aspects of RNA metabolism.
195
196 The FA model decomposed the 55 x 13,299 normalized RBP binding matrix into a 55 x 10 factor
197 loading matrix (representing the strength of the dependence of each of the 55 RBP target RNA
198 binding pattern on each of the 10 *factors*), a 13,299 x 10 factor score coefficient matrix
199 (representing the dependence between the binding of the 13,299 target RNA encoding gene and
200 each of the 10 *factors*), and residual error (Supplemental Fig 3a and methods). Cumulatively, the
201 FA model explained ~60% of the variance in the observed data. The remaining unexplained
202 variance was expected due to the challenges of integrating data sets of varying depth and quality,
203 in spite of our efforts to control these aspects. The communality, which is the amount of variance
204 explained by the model for each RBP-binding variable, varied drastically for all 55 RBPs; the
205 model explained at least 80% of the variance in enrichment scores for 12 RBPs, and at least 50%
206 of the variance in enrichment scores for 30 RBPs (Supplemental Figure 3b). RBPs with lower
207 communality often coincided with shallow depth of their PAR-CLIP libraries.
208
209 The FA model also uncovered interesting parallels between the similarity in the binding of target
210 RNA encoding genes and the target annotation category preferences (from Figure 1a). We
211 observed that individual *factors* contained RBPs that preferred binding to either mature (Factors
212 1, 3, 4, 5, 8) or precursor transcripts (Factors 2, 6), reflecting involvement in different stages of
213 RNA metabolism (Figure 3a). Furthermore, individual *factors* contained RBPs exhibiting similar
214 patterns of binding to specific regions of the mRNA (i.e., intron, coding, 3' UTR). Indeed, RBPs
215 from the same family, or known to regulate a specific aspect of RNA processing, had high

216 loadings for the same *factors*. For example, the ELAVL1 family members were associated with
217 Factor 1; the AGO1 family were associated with Factor 3; the IGF2BP1 family were associated
218 with Factor 4; the FMR1 family had were associated with Factor 5 and Factor 8; LINE-1
219 encoded proteins were associated with Factor 7. One of the unanticipated associations was that
220 of HNRNPC with Factor 2, which contained mainly cleavage and polyadenylation factors.
221 Interestingly, HNRNPC was shown to interact with U-rich sequences downstream of a viral
222 poly-adenylation signal nearly three decades ago (Wilusz et al., 1988), and more recently, to
223 repress cleavage and poly-adenylation in humans (Gruber et al., 2016). These examples highlight
224 the specific testable hypotheses generated by an integrative analysis that are not necessarily
225 obvious when examining a single RBP in isolation.
226
227 By clustering the factor score coefficients, i.e. the specific linear combination of RBP binding for
228 that target RNA, we identified target RNA encoding genes constituting putative regulatory
229 modules associated with a given *factor*. Therefore, each regulatory module was associated with
230 an RBP component (the subset of RBPs exhibiting similar binding pattern) and a RNA
231 component (the subsets of target RNA encoding genes bound by those RBPs). These regulatory
232 modules did not imply physical interactions between RBPs; rather, it identified RBPs that may
233 cooperate in controlling RNA metabolism for specific subsets of RNA targets, possibly across
234 cellular compartments. Almost a quarter of the target RNA encoding genes (3,180/13,299) were
235 assigned to regulatory modules by exhibiting high factor score coefficients for a single *factor*
236 (Supplemental figure 3c). We did not identify target RNA encoding genes with high factor score
237 coefficients for Factor 9 or 10. The remaining target RNA encoding genes did not exhibit high
238 factor score coefficients for any specific *factor* in our analysis, suggesting that the targets were

239 either not bound by specific combinations of these RBPs, bound broadly by all RBPs, or not
240 bound by the subset of RBPs in the analysis. As such, we labeled this target RNA encoding gene
241 category as "non-specific". The RNA regulatory modules encoding genes were enriched for
242 different GO categories. Factor 1 RNA regulatory modules were enriched for 'AU-rich element
243 binding' and Factor 3 RNA regulatory modules were enriched for 'gene silencing by miRNA';
244 AU-rich RBPs and AGO proteins were strongly associated with Factor 1 and Factor 3,
245 respectively. This was consistent with the recurrent observation that RBPs target the mRNAs
246 encoding themselves (Pullmann et al., 2007; Tenenbaum et al., 2000). In turn, the RNAs
247 encoding "non-specific" genes contained ribosomal proteins and mitochondrial electron-
248 transport proteins.
249

250 **RNA regulatory modules underlie distinct patterns of RNA metabolism**

251 In order to test the functional relevance of these RNA regulatory modules, we reasoned that
252 perturbation (change of protein abundance or activity) of an RBP will lead to pronounced effects
253 only for the RNA regulatory modules assigned to the specific *factor(s)* that RBP is associated
254 with. We examined mature and precursor RNA expression changes induced by siRNA
255 knockdown of ELAVL1 (Kishore et al., 2011). ELAVL1 was strongly associated with both
256 Factor 1 and Factor 2, which exhibited RNA targeting patterns for mature or precursor RNAs,
257 respectively. Concordantly, Factor 1 associated RNA regulatory modules, but not Factor 2 RNA
258 regulatory modules, exhibited ELAVL1-dependent stabilization of mature RNA (Figure 4a).
259 Likewise, Factor 2 RNA regulatory modules exhibited a more pronounced ELAVL1-dependent
260 stabilization of precursor RNA than Factor 1 RNA regulatory modules (Figure 4b). Each human
261 ELAV1 family protein contains three RRM domains (>90% sequence identity), but the hinge

262 region between the second and third RRM of ELAVL1 contains a shuttling sequence responsible
263 for its nuclear localization (Fan and Steitz, 1998). Due to the lack of this shuttling sequence,
264 ELAVL2/3/4 are predominantly cytoplasmic and were strongly associated with Factor 1, but not
265 Factor 2. Taken together, the model was able to correctly identify and distinguish ELAVL1-
266 dependent stabilization of both precursor and mature RNA (Lebedeva et al., 2011; Mukherjee et
267 al., 2011).
268
269 We hypothesized that the subsets of RNAs assigned to the different regulatory module would
270 exhibit differences in RNA metabolism driven by the RBPs in the *factor* associated with the
271 regulatory module. Therefore, we compared six aspects of RNA metabolism previously
272 quantified in HEK293 cells (Mukherjee et al., 2017), for each of the RNA regulatory modules
273 associated with each of the *factors*. The *factor*-associated RNA regulatory modules exhibited
274 very distinct RNA metabolic profiles compared to each other and to non-specific category
275 (Figure 4c, Supplemental Figure 4a). Factor 2 RNA regulatory modules, which was the only
276 factor associated with RBPs binding to precursor mRNA and lncRNA, had low processing rates,
277 high degradation rates and their encoded RNAs were preferentially localized in the nucleus
278 versus the cytoplasm. Factor 2 RNA regulatory modules were strongly enriched for lncRNAs
279 (Figure 4d). Indeed, these genes strongly overlapped with a set of lncRNAs likely to be
280 functional (Supplemental figure 4b) (Mukherjee et al., 2017).
281
282 We also examined regulatory differences in RNA metabolism for genes associated with
283 cytoplasm-enriched factors. For example, factor 1 RNA regulatory modules were more stable
284 than Factor 3 RNA regulatory modules (Figure 4c). Factor 1 was strongly associated with

285 ELAVL1 family proteins, which stabilize target mRNAs. Factor 3 was strongly associated with
286 AGO1 family proteins, which execute miRNA-mediated degradation of target mRNAs.
287 Additionally, Factor 4 RNA regulatory modules, which are bound by IGF2BP1 family proteins,
288 were highly synthesized, processed, stabilized, and translated (Figure 4c). The RNA targets of
289 IGF2BP1 family RBPs were strongly localized to the ER (Supplemental Figure 4c) (Jønson et
290 al., 2007), which is also consistent with the proposed role of IGF2BP1 family proteins for RNA
291 localization and translation (Farina et al., 2003; Nielsen et al., 2001). Although correlative, these
292 results indicate that different RBP binding patterns beget different consequences for RNA
293 metabolism.
294
295 Specific RNA regulatory modules also exhibited preferential localization to processing bodies
296 (P-bodies), which are cytoplasmic granules associated with translational repression (Sheth and
297 Parker, 2003). Namely, Factor 3 RNA regulatory modules, which were strongly associated with
298 the AGO1 family, were the most strongly enriched for localizing to P-bodies according to a
299 recent study characterizing the transcriptome and proteome of P-bodies, and the AGO2 protein
300 itself was 90-fold enriched (Hubstenberger et al., 2017). Similarly, Factor 5 RNA regulatory
301 modules, which were strongly associated with the FMR1 family, were also enriched for
302 localizing in P-bodies, along with the FMR1 protein (16-fold enriched). In contrast, the non-
303 specific category was depleted from P-bodies.
304
305 Fine-tuning of gene expression has been postulated to be an important function of post-
306 transcriptional regulation by RBP and miRNAs. Therefore, we examined the cell-to-cell
307 variability in gene expression across 25 individual HEK293 cells with respect to the RNA

308 regulatory modules. The single-cell RNA-seq data was very deeply sequenced and generated
309 using the massively parallel single-cell RNA-sequencing (MARS-Seq) protocol (Guillaumet-
310 Adkins et al., 2017). Most RNA regulatory modules exhibited lower expression variability than
311 the non-specific category (Figure 4e). In particular, Factor 4 RNA regulatory modules exhibited
312 the lowest variation and highest median expression across the 25 cells (Supplemental Figure 4d).
313 These results supported the broad notion that post-transcriptional gene regulation generally
314 confers robustness and fine-tuning of gene expression.
315

316 **Conclusion**

317 Our study presents a curation of existing datasets, followed by systematic analysis of high-
318 quality and high-resolution RBP-RNA interaction data. We focused on the RBPs that
319 preferentially bound to mRNA and lncRNA and examined their sequence specificity and
320 sequence motif preferences. Our survey of the RBP regulatory landscape identified the most
321 prevalent subsets of RNAs targeted by a specific subset of RBPs, which we refer to as RNA
322 regulatory modules.
323

324 We utilized high quality PAR-CLIP datasets for which the immunoprecipitation was generally
325 comparable due to fact most RBPs were FLAG-tagged. Nevertheless, several caveats associated
326 with the interpretation of this analysis need to be pointed out. Despite several measures of quality
327 control to decide which datasets to include in our analysis, the libraries varied greatly in depth,
328 quality, digestion biases and potentially other confounding variables with respect to the protocol.
329 The FA model quantitatively assessed the degree to which we could explain the full complement
330 of RBP-RNA target binding patterns. These confounders undoubtedly contributed to the ~40%

331 of variance not explained by the FA model. In comparison, the ENCODE eCLIP datasets (Van
332 Nostrand et al., 2016) are likely to suffer from different confounders: they were generated using
333 one consistent experimental protocol but used antibodies against endogenous proteins expressed
334 at varying levels, and for which IP efficiency can vary greatly in spite of the quality control
335 performed (Sundararaman et al., 2016). Essentially, this represents the trade-offs in experimental
336 design between analyzing the endogenous protein compared to an epitope-tagged protein.
337 Modifying the genomic loci of the protein to engineer an endogenous epitope tagged RBP is
338 a very promising strategy.
339
340 Assuming the RBPs investigated here are a representative sample of the ~1,542 RBPs encoded in
341 the human genome, there may be an astounding number of RBPs with substantial primary
342 sequence specificity. However, the degree of sequence specificity is determined by the nature of
343 the RBP-RNA interaction, which can be quite extensive and specific, as in the case of Pumilio,
344 or minimal and non-sequence specific, as in the case of an RNA-helicase. An interesting
345 exception were the A-rich sequences enriched by UPF1, which is an RNA helicase and therefore
346 unlikely to exhibit strong sequence specificity. One possible explanation is that such sequences
347 may represent pre-mature polyA tail recognition involved in aspects of ribosome quality control
348 demonstrated in yeast (Koutmou et al., 2015) and human cells (Garzia et al., 2017a). Likewise,
349 more examples of unanticipated sequence enrichments may shed light on novel RNA regulatory
350 mechanisms.
351
352 Our FA model was able to identify distinct RBP-RNA target regulatory modules. At the very
353 minimum, 25% of target RNA encoding genes were assigned to RNA regulatory modules. This

354 is very likely an underestimation due to noisy data and a biased, far from complete sampling of
355 RBPs. However, there is likely to be a subset of genes for which post-transcriptional gene
356 regulation indeed plays a negligible role, at least in HEK293 cells. Furthermore, a small number
357 of RBPs in our analysis are not endogenously expressed in HEK293 and their natural expression
358 is tissue-specific and/or context-dependent. The approach presented here can scale to binding
359 data for all ~700 RBPs experimentally shown to be associated with poly-adenylated RNA in
360 HEK293 cells or even ~1,542 known RBPs (Baltz et al., 2012).
361
362 The RNA regulatory modules exhibited different patterns of RNA processing, degradation,
363 localization, and translation. We speculate that these differences in RNA metabolism were driven
364 by individual RBPs or the combination of RBPs associated with that regulatory module. This
365 was supported by the response of specific RNA regulatory modules to ELAVL1 knockdown
366 (Figure 4A, B). Additionally, the RNA regulatory modules encoded functionally related proteins
367 and similarly localized proteins. The enrichments were for proteins with similar molecular
368 functions or multi-component complexes rather than signaling pathways (Supplemental Fig 3b).
369 Altogether, these lines of evidence provide support for the coordinate regulation of 'functionally
370 coherent' RNA regulatory modules as proposed by the post-transcriptional operon/regulon model
371 (Keene, 2007). The ultimate test of this model would involve manipulating specific combinations
372 of binding sites and RBPs. Our study provides the rationale for such experiments, which
373 unfortunately remain technically challenging.
374
375 Our observations have important implications for RBP-RNA regulatory networks and their
376 importance in gene expression. The mRNA targets within specific regulatory modules encoded

377 the RBP themselves, a generalization of a commonly made observation that RBPs bind to the
378 mRNAs encoding them (Mesarovic et al., 2004). Our analysis lends support for this frequently
379 observed potential auto-regulatory feedback. These feedback loops may in fact buffer the
380 expression range of the targeted mRNAs, including those of the RBP. In this context, the
381 observation that the RNA regulatory modules exhibited lower cell-to-cell gene expression
382 variance, provides more evidence for the importance of post-transcriptional regulation in
383 buffering transcriptional noise (Bahar Halpern et al., 2015; Battich et al., 2015). Systematic
384 perturbation of individual and combinations of RBPs will be quite powerful in revealing
385 fundamental properties of RNA regulatory networks such as auto-regulatory feedback and
386 buffering.
387
388 The binding preference and targets of the vast majority of human RBPs remains unknown. The
389 insights gained from this study demonstrate the value of large-scale efforts by ENCODE and
390 others in the community to globally identify RBP binding sites. Of the 64 RBPs in this study, 44
391 were not represented in the ENCODE cell lines. Cumulatively these efforts interrogate ~10% of
392 human RBPs with known RNA-binding domains. Thus, these two large scale efforts offer the
393 potential to complement one another in our continuing attempts to understanding RBP-RNA
394 regulatory networks, for which we have only glimpsed the tip of the iceberg.
395

396 **Acknowledgements**

397 U.O. and T.T. acknowledge support from an award from the US National Institutes of Health
398 (R01-GM104962). M.S. was supported by the ETIUDA scholarship # 2014/12/T/NZ1/00497
399 from National Science Center, Poland. N.M. acknowledges support from EU Marie Curie IIF,
400 and the RNA Bioscience Initiative for startup funds.

401 **Declaration of Interests**

402 T.T. is cofounder and advisor to Alnylam Pharmaceuticals.

403 **Author Contributions**

404 Conceptualization, N.M., T.T, and U.O.; Methodology, N.M., H.W, M.S., M.G, A.G, A.M., T.T,
405 U.O; Investigation, N.M., H.W, M.G, A.G, A.M., T.F, J.I.H, K.A; Formal Analysis, N.M., H.W,
406 M.S., M.G, A.G, A.M. ; Writing – Original Draft, N.M., T.T, and U.O.; Writing – Review &
407 Editing, N.M., S.V. A.G. M.S., T.T, and U.O.; Funding Acquisition, N.M., T.T, and U.O.;
408 Resources, N.M., T.T, and U.O.Supervision, N.M., T.T, and U.O.

409

410 **STAR Methods**

411 **Processing, filtering, and quality control of PAR-CLIP libraries**

412 Each PAR-CLIP library was subject to two rounds of quality control. First, all PAR-CLIP
413 libraries generated in HEK293 cells were subject to the quality control pipeline PAR-CLIP Suite
414 v1.0 (https://rnaworld.rockefeller.edu/PARCLIP_suite/). Using raw Illumina sequencing data,
415 this pipeline identified the predominant target RNA category or categories for each RBP and
416 provided the T-to-C conversion frequency resolved by read length and RNA category

417 (Supplemental Fig 1). The mapped reads of each RNA category were resolved by error distance
418 0 (d0), error distance 1 (d1; split in T-to-C and d1 other than T-to-C), and error distance 2 (d2).
419 This process discriminated for each library true target RNA categories from non-crosslinked
420 background RNA categories populated by fragments of abundant cellular RNAs. In order to
421 disqualify experiments comprising too many non-crosslinked RBP-specifically bound RNAs or
422 co-purified non-crosslinked background RNAs, we pursued only datasets which collect at least
423 10,000 redundant d1 reads ≥ 20 nt in at least one of major RNA annotation categories with d1(T-
424 to-C)/(d0 + d1) ≥ 30%, and d1(T-to-C)/(d1-total) ≥ 65%.
425 For the libraries passing the first threshold, we defined and annotated binding sites using
426 PARpipe, which is a pipeline wrapper for PARalyzer (Corcoran et al., 2011; Mukherjee et al.,
427 2014). The threshold for additional filtering were determined by comparisons with the reference
428 library (Friedersdorf and Keene, 2014). This reference library was generated using a modified
429 PAR-CLIP protocol in which there was no immunoprecipitation and the addition of an rRNA
430 depletion step after proteinase K digestion, followed by a partial digestion using RNase T1. We
431 required libraries had to have an average fraction T-to-C over remaining reads greater than 0.32
432 (the average fraction T-to-C over remaining reads greater of the reference library), an average
433 conversion specificity greater than 0, more than 20000 aligned reads, not be digested only with
434 micrococcal nuclease, a redundant read copy fraction less than .98 (Supplemental Fig 1b,c and
435 Sup Table 1). For RBPs with three or more libraries, we removed outlier based on correlation of
436 6-mer frequency calculated from PARalyzer-utilized reads.
437

438 **Annotation category preference and positional analysis of binding density**

439 For calculating the annotation category preference, we calculated the difference in the fraction of
440 T-to-C reads per annotation category between each RBP library and the reference library. For
441 example, if the fraction of miRNA annotated reads with T-to-C transitions in a specific RBP
442 library was 0.20 compared to 0.05 in the reference library, the miRNA preference value for this
443 specific RBP is 0.15. For the positional binding analysis, we selected genes (n=15120) using
444 GENCODE v19 as annotation based on our earlier work on HEK293 RNA processing and
445 turnover dynamics (Mukherjee et al., 2017). Isoform expression was calculated using RSEM (Li
446 and Dewey, 2011). For each gene, we selected the transcript isoform with the highest isoform
447 percentage or chose one randomly in case of ties (n=8298). The list of selected transcript
448 isoforms was used to calculate the median 5' UTR, CDS and 3' UTR length proportions (5'
449 UTR=0.06, CDS=0.53, 3' UTR=0.41) using R Bioconductor packages GenomicFeatures and
450 GenomicRanges. For regions downstream annotated transcription ends (TES) and adjacent to
451 splice sites, we chose windows of fixed sizes (TES 500nt, 5' and 3' splice sites 250nt each). We
452 generated coverage tracks from the PARalyzer output alignment files and intersected those with
453 the filtered transcripts. Each annotation category was binned according to its relative coverage
454 averaged according to each bin. For intronic coverage, we averaged across all introns per gene,
455 given a minimal intron length of 500nt. All bins were stitched to one continuous track per
456 transcript. Altogether 6632 intron containing transcripts showed coverage in at least one
457 PARCLIP library. For each library, we required transcripts to have a minimal coverage
458 maximum of > 2. For each transcript, we scaled the binned coverage dividing by its maximal
459 coverage (min-to-1 scaling) to emphasize spatial patterns independent from transcript expression
460 levels. Replicate RBP PARCLIP libraries were combined at this point. Transcripts targeted in

more than one replicate library were aggregated using the average of their binned coverage. RBPs with less than 50 filtered target transcripts (after aggregation) were not considered. Next, we split transcript coverage in two parts, separating 5' UTR to TES regions and intronic regions. To generate the scaled meta coverage across all targeted transcripts per RBP, we used the heatMeta function from the Genomation package. For the 5'UTR to TES, we scaled each RBP meta-coverage track independent of other RBPs. For each RBP, we subtracted the scaled meta coverage of PARCLIP reference library (Friedersdorf and Keene, 2014). For intronic sequences, we scaled each RBP relative to all other RBPs to highlight RBPs with more substantial intronic binding patterns. Finally, we visualized the density using pheatmap.

Sequence analysis

We calculated 6-mer frequencies with Jellyfish from all reads that generated a PARalyzer binding site for each library. For each RBP, we selected the library with the lowest percent of duplicated sequences (see supplemental table 1) to serve as a representative library for the sequence analysis and factor analysis. For each RBP, we counted the number of 6-mers with a frequency of x or higher, where x was from 1/4096 to 1/4. To evaluate the 6-mers enriched by a given RBP relative to the reference library, we regressed the RBP 6-mer frequency against the the reference library 6-mer frequency and collected the residuals (the unexplained variance). Next, identified all 6-mers that were found as the top 5 enriched over the reference library for any of the analyzed RBPs. We clustered the enrichment scores for the 6-mers across all RBPs and generated a heatmap using the ‘aheatmap’ function in NMF R package. We ran SSMART using all binding sites found in mRNA-derived annotation categories ranked by the library size normalized enrichment over the reference library.

484 **Factor analysis**

485 For each site identified we calculated a library size normalized enrichment compared to the the
486 reference library. We calculated the sum of all enrichment scores for all sites annotated as
487 mRNA and lncRNA. Next, we normalized for expression levels (collected the residuals) to
488 create the final matrix of values. The number of factors, 10, was determined using the majority
489 result of numerous methods to estimate the number of factors. Clustering of the score matrix was
490 performed using the most stable results from numerous iterations of k-means clustering.
491

492 **Gene ontology analysis**

493 Multiple-test corrected gene ontology enrichment values were calculated using the TOPGO R
494 package. For each set of genes, we used all 13,299 genes in the factor analysis as the background
495 or gene universe. Enrichment was calculated using the 'parent-child' approach on the top 100
496 enriched terms. This metric accounts for the hierarchical organization of gene ontology terms to
497 minimize false-positive enrichments. We performed a Bonferonni multiple test correction on the
498 enrichment p-values.
499

500 **Premature and mature RNA quantification**

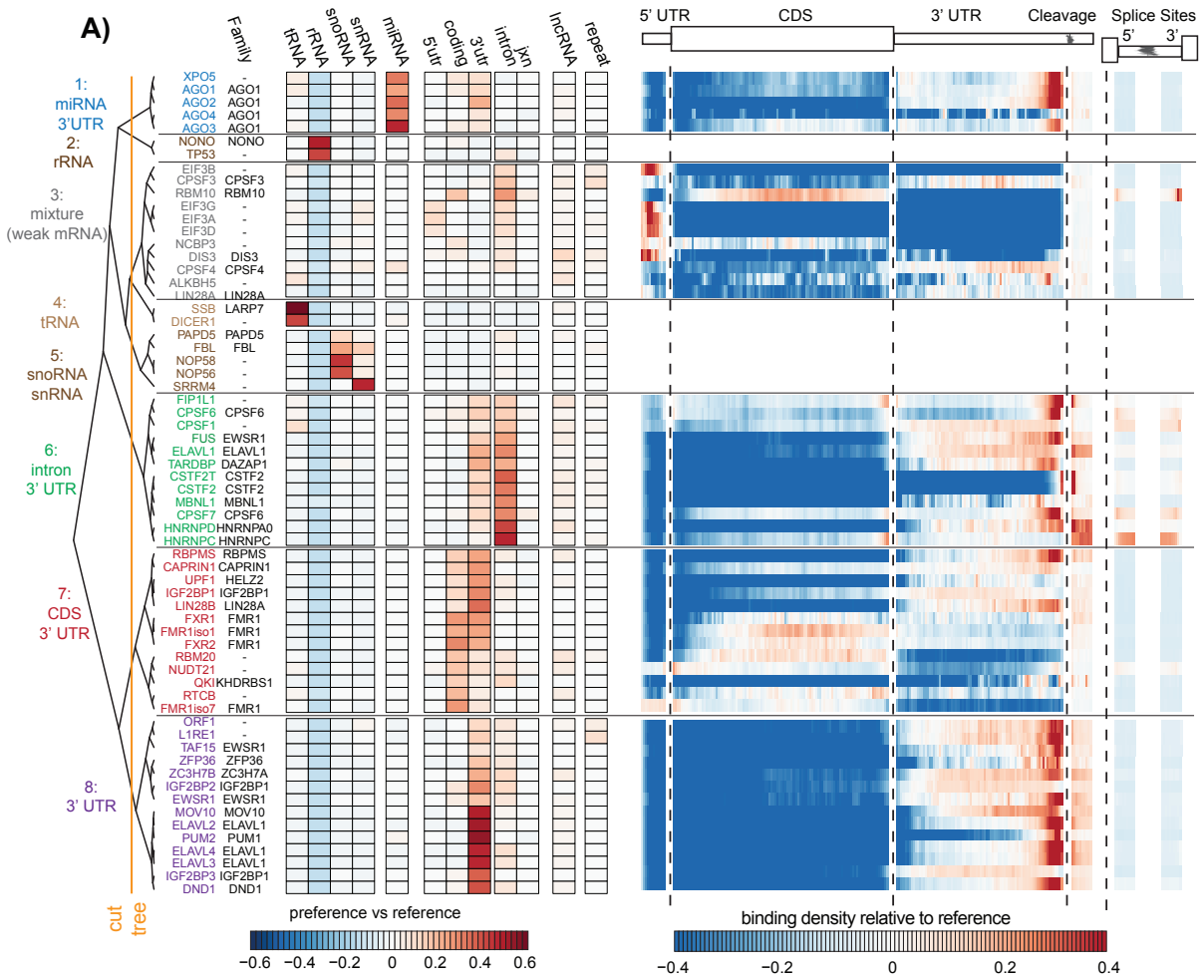
501 Mature- and premature-transcript expression, transcripts per million (TPM), was quantified with
502 RSEMv1.2.11 (http://deweylab.biostat.wisc.edu/rsem/src/rsem-1.2.11.tar.gz) as described
503 previously (Mukherjee et al., 2017). Briefly, for each gene we included an additional isoform
504 corresponding to the sequence of the full gene locus. Specifically, we modified the
505 GENCODEv19 gtf and used this as the input for the 'rsem-prepare-reference' function to

506 generate a modified index used for quantification. For each gene, we calculated the expression of
507 'mature' RNA as the sum of all isoforms for that gene excluding the 'primary' transcript. For
508 intronless genes, premature and mature expression values were summed. We performed this
509 analysis on the ELAVL1 knockdown RNA-seq experiments (Kishore et al., 2011).
510

511 **Cell-to-cell expression variability**

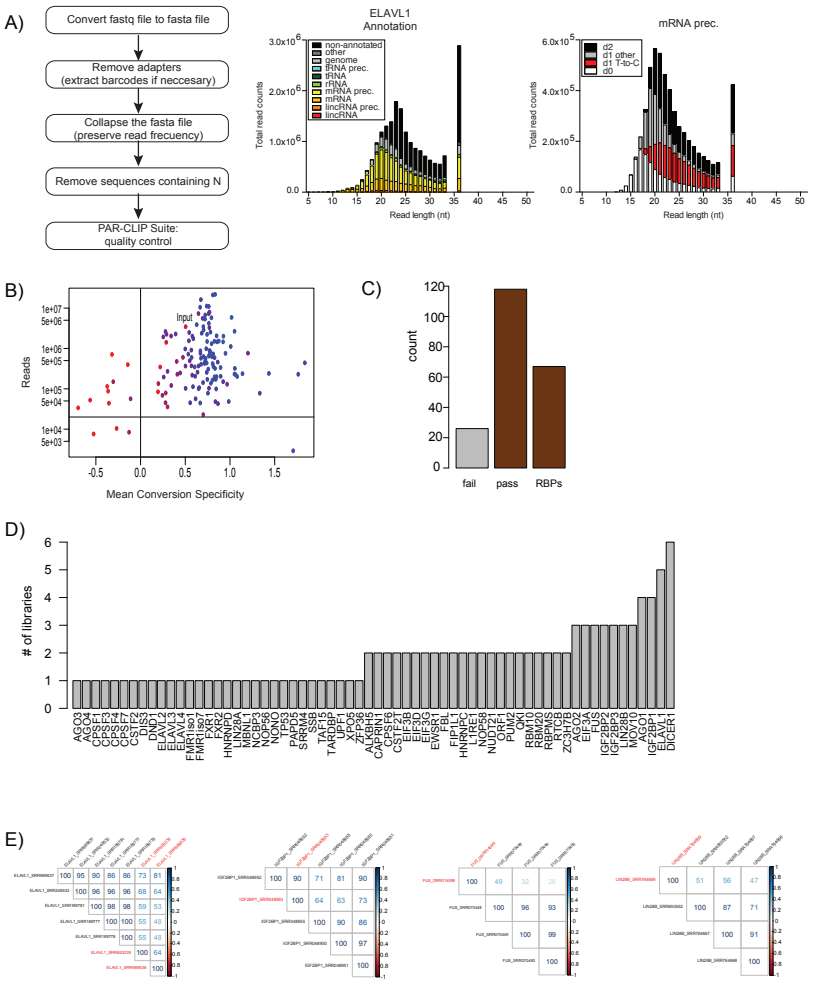
512 RNA-seq gene expression data for 25 individual HEK293 cells were downloaded from
513 (Guillaumet-Adkins et al., 2017). We calculated the coefficient of variation (100*standard
514 deviation/mean) for each gene across all 25 cells.

515 **Figures**



516 **Figure 1. RBP analyzed and binding preferences by RNA category.** A) Heatmap of reference
517 normalized annotation category preference for each RBP clustered into 8 branches by color
518 (left). The heatmap represents the difference in the proportion of sites for a given annotation
519 category in the RBP library versus the reference library. Heatmap of the reference library
520 normalized relative positional binding preference of the 55 RBPs with enriched binding in at
521 least one mRNA-relevant annotation category per branch (right). RBP-specific binding
522 preferences were averaged across selected transcripts (see methods). The relative spatial
523 proportion of 5'UTR, coding regions and 3'UTR were averaged across all selected transcript

524 isoforms. For TES (regions beyond transcription end site), 5' splice site, and 3' splice site, we
525 chose fixed windows (250nt for TES and 500nt for splice sites). For each RBP, meta-coverage
526 was scaled between 5'UTR to TES. The 5' and 3' intronic splice site coverage was scaled
527 separately from other regions but relative to each other.
528



Supplemental Figure 1. QC filtering of libraries. A) Description of PAR-CLIP suite to assess library quality control per annotation category (left). Example of number of reads mapping to each RNA category with up to 2 mismatches resolved by length of adapter-extracted sequence reads for an ELAVL1 library (middle). Sequencing read composition of the most abundant RNA category fir the ELAVL1 library. Reads were assigned as d0 (white), d1 T-to-C (red), d1 other than T-to-C, (light gray), and d2 (black) (right). B) Libraries had to have > 20,000 aligned reads and a mean conversion specificity > 0, and a higher mean T-to-C fraction than the the reference library (red lower, blue higher). C) Number of libraries analyzed and their quality control status. D) Count of libraries passing QC per RBP. E) Examples of outlier library removal (libraries labeled with red text were removed) based on correlation of read 6-mer frequency for RBPs with 3 or more libraries.

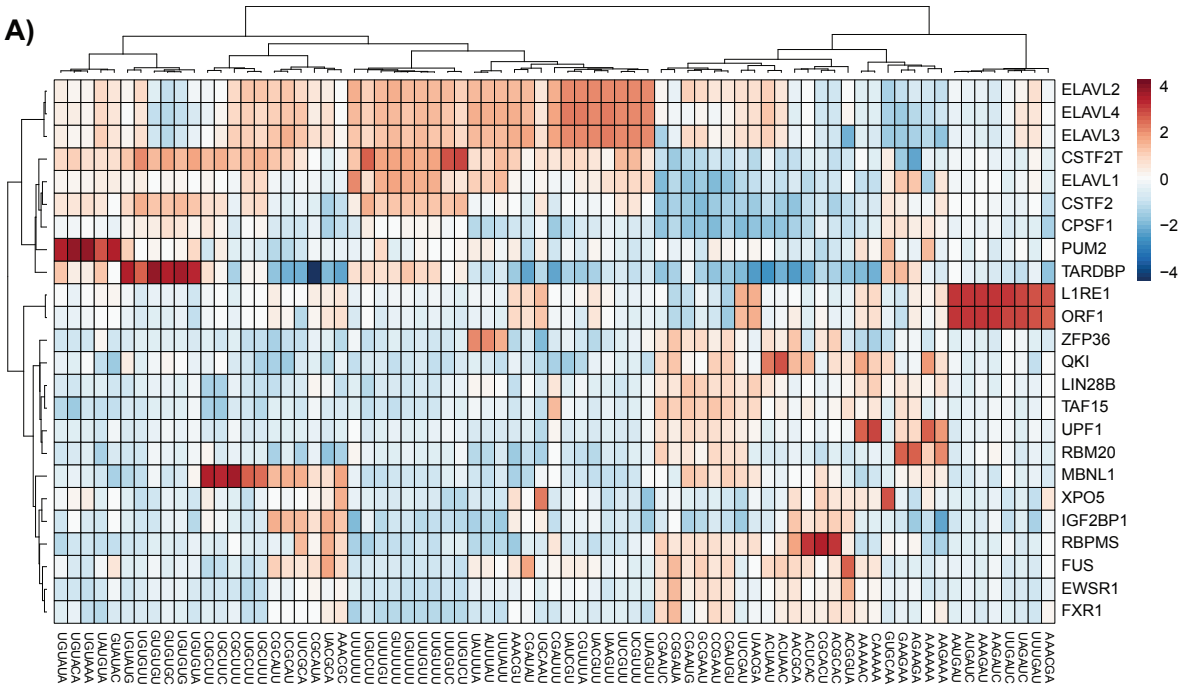
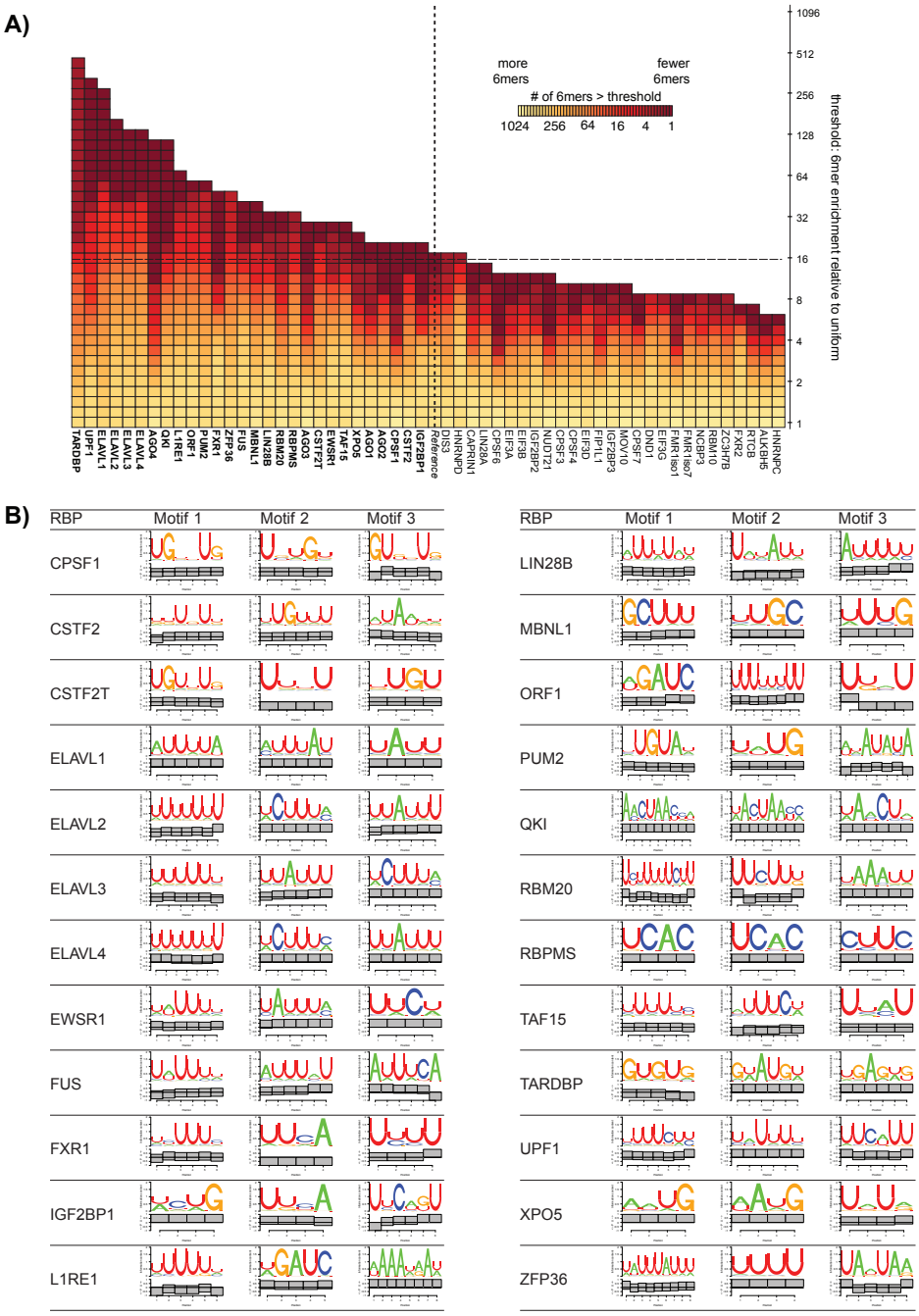
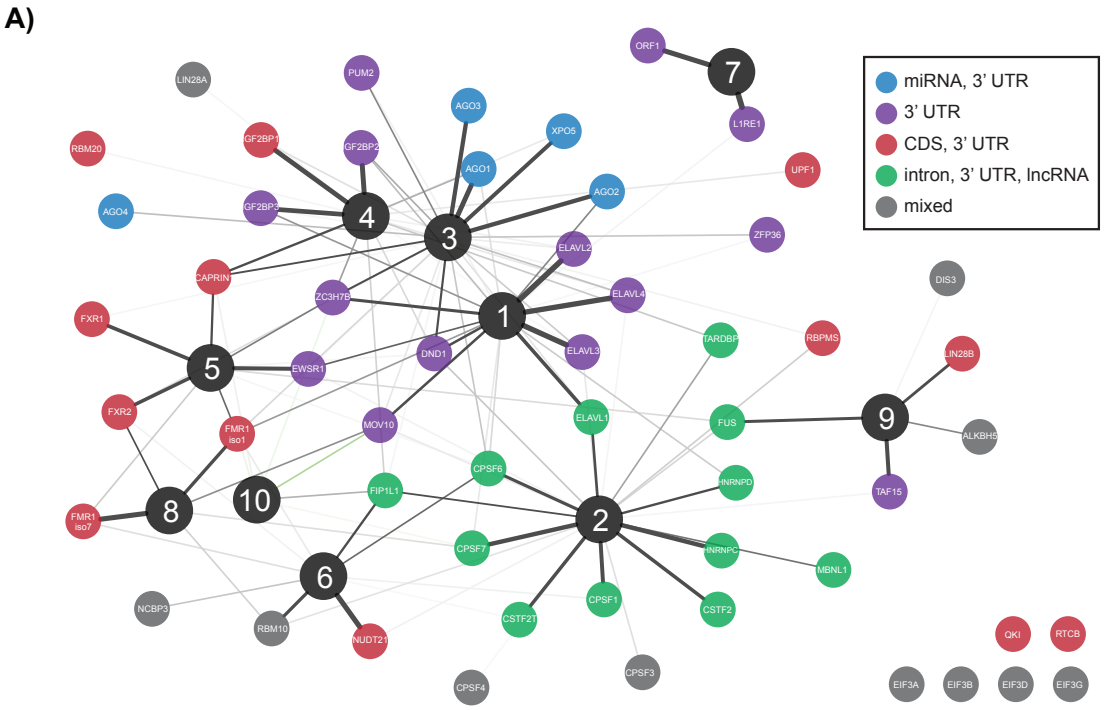


Figure 2. RBP binding sequence specificity and elements. A) Heatmap of reference normalized 6-mer enrichment for top 5 enriched 6-mers for each RBP in the set of RBPs exhibiting more sequence specificity than the reference.



Supplemental Figure 2. Grouping RBPs by sequence specificity. A) Heatmap of the number of 6-mers enriched per RBP at different specificity thresholds. The color scale represents the $\log_2$ [number of 6-mers] that are enriched at a given threshold (y-axis). The thresholds are represented as $\log_2$ [6-mer frequency]. There are 4096 different 6-mers and if they were uniformly present this would represent a value of -12 $=\log_2$ [1/4096]. The horizontal dashed lines at -8, represents 16-fold enrichment over a uniform background. For reference, the vertical dashed lines indicate the behavior of the reference library. B) Top 3 SSMART motif results using all binding sites found in mRNA-derived annotation categories ranked by the library size normalized enrichment over reference library.

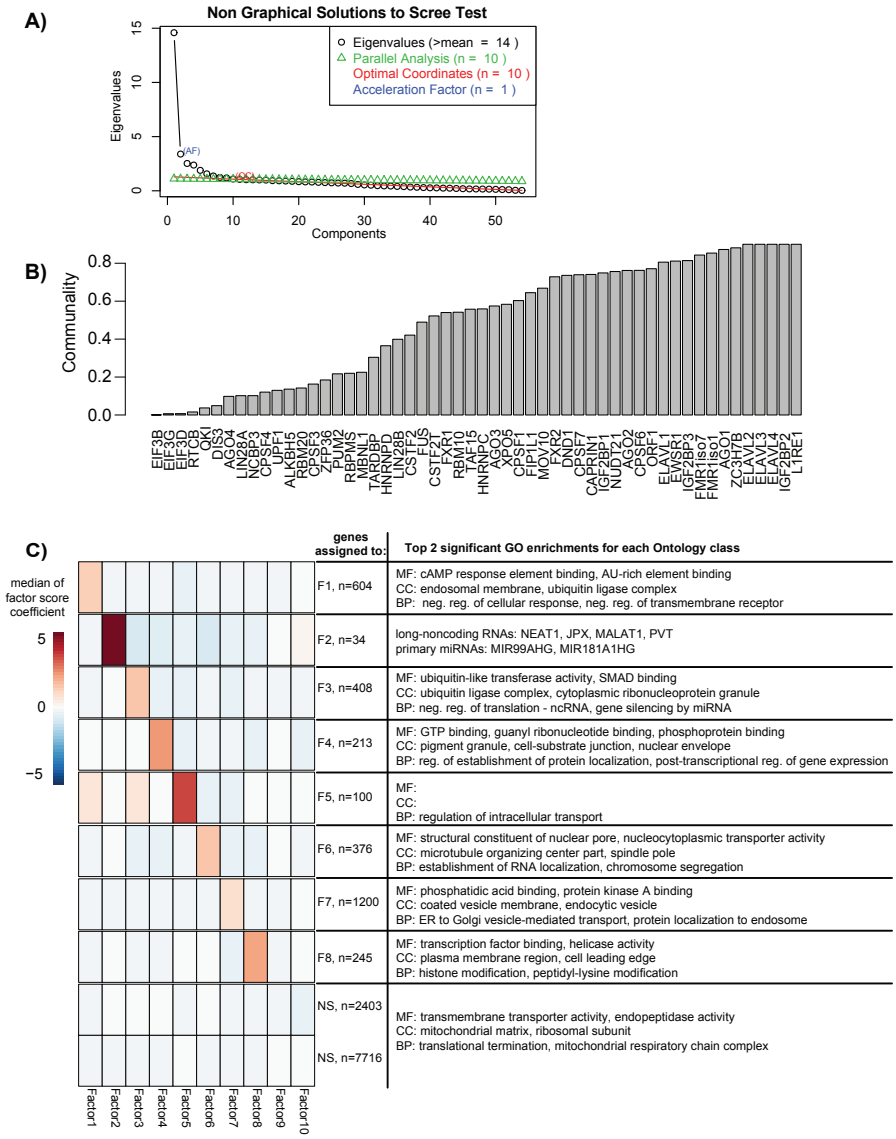


578

579 **Figure 3. RNA regulatory modules.** A) Factor analysis of target RNA encoding genes binding
580 normalized by the reference library and expression for the 55 RBPs binding to mRNAs and
581 lncRNAs for 13,299 genes (see 'factor analysis' section in methods for details). Spring-
582 embedded graph of the factor loading matrix, indicating the association between each of the 55
583 RBPs and one of the 10 factors. Nodes color-coded by RNA annotation category preference
584 cluster membership from figure 1. Edge width scales with factor loadings (thicker edge = higher
585 factor loading = stronger association). Only edges with a factor loading > 0.2 (positive values in
586 black) or < -.2 (negative values in green) depicted.

587

Supplemental Figure 3. Factor analysis model selection and performance. A) Plot of eigenvalues versus number of factors to determine the optimal number of factors using four methods (different colors). B) Barplot of the communality, or the variance in a given RBP cumulatively explained by the all factors. C) Heatmap of the median factor score coefficient value for all genes that clustered together. The number of genes assigned to a specific factor and the top two most significant enriched GO annotations for each ontology class: molecular function (MF), cellular component (CC), and biological process (BP).



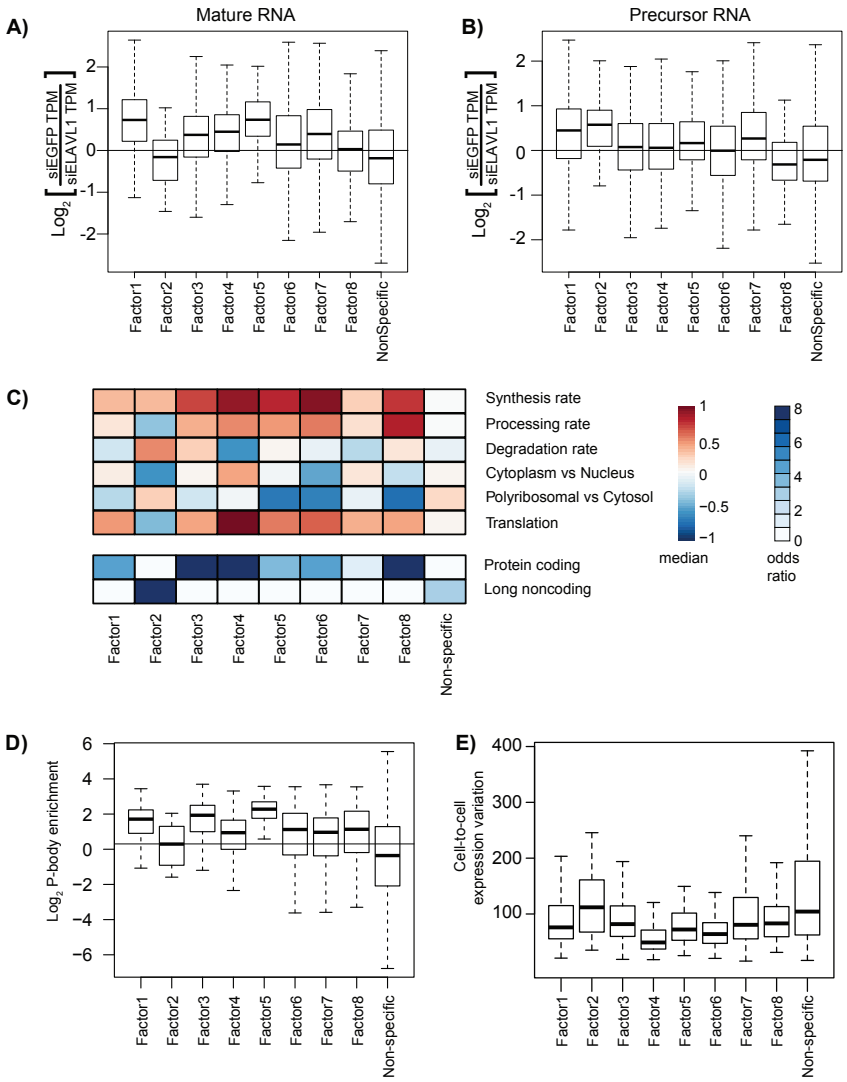
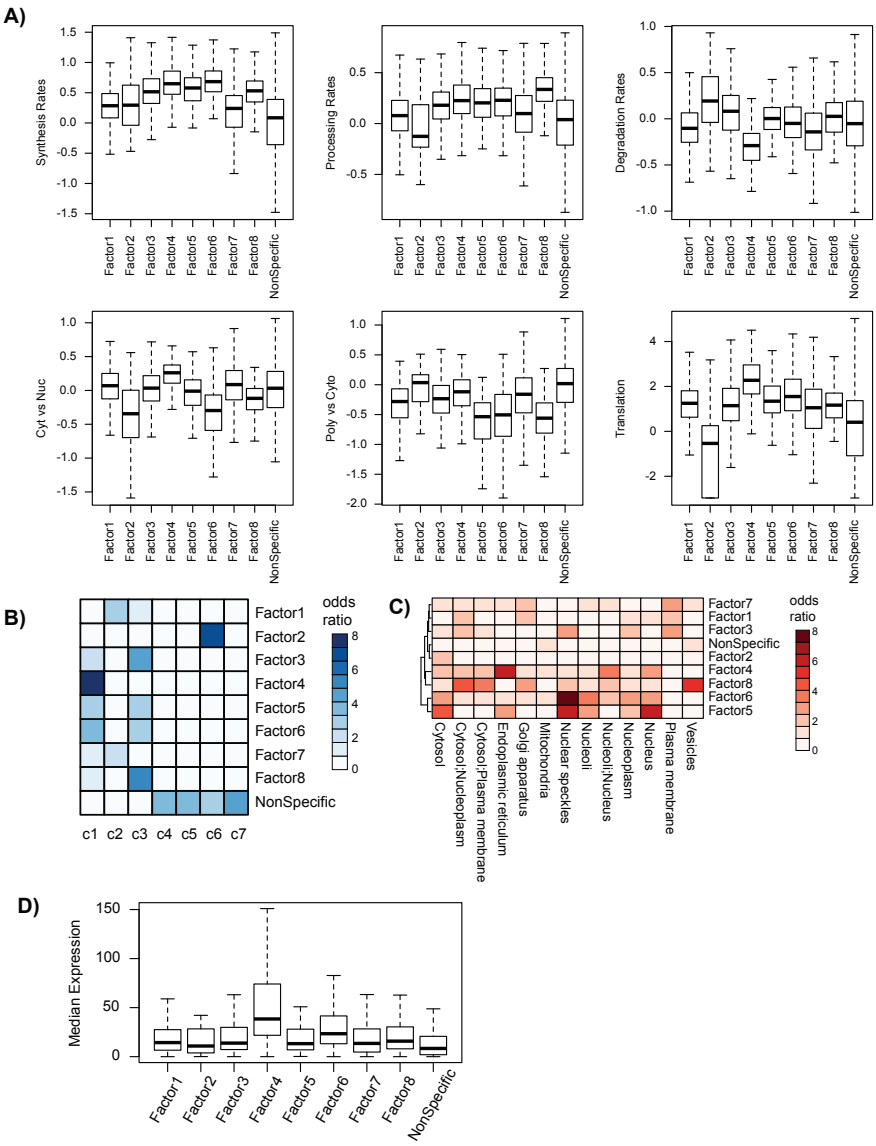


Figure 4. Functional characterization of RNA regulatory modules. A) The difference in either A) primary or B) mature RNA expression (transcripts per million) upon ELAVL1 knockdown by siRNA treatment (y-axis), specifically the $\log_2$[siRNA EGFP TPM]-$\log_2$[siRNA ELAVL1 TPM], for each gene set. C) Heatmap of the median value of synthesis rate, processing rates, degradation rates, cytoplasmic versus nuclear localization, polyribosomal versus cytoplasmic localization, and translational status from ribosome profiling data for each gene set (top). Heatmap of the odds-ratio of the overlap between factor associated gene sets with annotation (bottom). D) Box-and-whisker plot for each gene set of the enrichment in P-bodies. E) Box-and-whisker plot for each gene set of the coefficient of variation across 25 individual HEK293 cells.



631 **Supplemental Figure 4. RNA metabolism profiles for factor-associated gene sets.** A) Box-and-whisker plot for each gene set of the synthesis rates, processing rates, degradation rates, cytoplasmic versus nuclear localization (Cyt vs Nuc), polyribosomal versus cytoplasmic localization (Poly vs Cyt), and translational status from ribosome profiling data. B) Heatmap of the odds-ratio of the overlap between factor associated gene sets with RNA categories based on similar metabolic profiles from (Mukherjee et al., 2017). C) Heatmap of the odds-ratio of the overlap between factor associated gene sets and protein localization annotation. D) Box-and-whisker plot for each gene set of the median expression across 25 HEK293 cells.

References

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., Itzkovitz, S., 2015. Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep 13, 2653–2662. doi:10.1016/j.celrep.2015.11.036

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., Landthaler, M., 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell 46, 674–690. doi:10.1016/j.molcel.2012.05.021

Battich, N., Stoeger, T., Pelkmans, L., 2015. Control of Transcript Variability in Single Mammalian Cells. Cell 163, 1596–1610. doi:10.1016/j.cell.2015.11.018

Cooper, T.A., Wan, L., Dreyfuss, G., 2009. RNA and Disease. Cell 136, 777–793. doi:10.1016/j.cell.2009.02.011

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., Ohler, U., 2011. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. 12, R79. doi:10.1186/gb-2011-12-8-r79

Fan, X.C., Steitz, J.A., 1998. HNS, a nuclear-cytoplasmic shuttling sequence in HuR. Proc. Natl. Acad. Sci. U.S.A. 95, 15293–15298.

Farina, K.L., Hüttelmaier, S., Musunuru, K., Darnell, R., Singer, R.H., 2003. Two ZBP1 KH domains facilitate β-actin mRNA localization, granule formation, and cytoskeletal attachment. The Journal of Cell Biology 160, 77–87. doi:10.1083/jcb.200206003

Fredericks, A.M., Cygan, K.J., Brown, B.A., Fairbrother, W.G., 2015. RNA-Binding Proteins: Splicing Factors and Disease. Biomolecules 5, 893–909. doi:10.3390/biom5020893

Friedersdorf, M.B., Keene, J.D., 2014. Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome Biol. 15, R2. doi:10.1186/gb-2014-15-1-r2

Garzia, A., Jafarnejad, S.M., Meyer, C., Chapat, C., Gogakos, T., Morozov, P., Amiri, M., Shapiro, M., Molina, H., Tuschl, T., Sonenberg, N., 2017a. The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. Nat Commun 8, 16056. doi:10.1038/ncomms16056

Garzia, A., Meyer, C., Morozov, P., Sajek, M., Tuschl, T., 2017b. Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. Methods 118-119, 24–40. doi:10.1016/j.ymeth.2016.10.007

Gerstberger, S., Hafner, M., Tuschl, T., 2014. A census of human RNA-binding proteins. Nature Reviews Genetics 15, 829–845. doi:10.1038/nrg3813

Gruber, A.J., Schmidt, R., Gruber, A.R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., Zavolan, M., 2016. A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. Genome Res. 26, 1145–1159. doi:10.1101/gr.202432.115

Guillaumet-Adkins, A., Rodríguez-Esteban, G., Mereu, E., Mendez-Lago, M., Jaitin, D.A., Villanueva, A., Vidal, A., Martinez-Marti, A., Felip, E., Vivancos, A., Keren-Shaul, H., Heath, S., Gut, M., Amit, I., Gut, I., Heyn, H., 2017. Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol. 18, 45. doi:10.1186/s13059-017-1171-9

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G.S., Dewell, S.,

Zavolan, M., Tuschl, T., 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141. doi:10.1016/j.cell.2010.03.009

Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.-B., Munier, A., Fradet, M., Daunesse, M., Bertrand, E., Pierron, G., Mozziconacci, J., Kress, M., Weil, D., 2017. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Mol. Cell 68, 144–157.e5. doi:10.1016/j.molcel.2017.09.003

Jankowsky, E., Harris, M.E., 2015. Specificity and nonspecificity in RNA-protein interactions. Nat. Rev. Mol. Cell Biol. 16, 533–544. doi:10.1038/nrm4032

Jønson, L., Vikesaa, J., Krogh, A., Nielsen, L.K., Hansen, T.V., Borup, R., Johnsen, A.H., Christiansen, J., Nielsen, F.C., 2007. Molecular composition of IMP1 ribonucleoprotein granules. Mol. Cell Proteomics 6, 798–811. doi:10.1074/mcp.M600346-MCP200

Keene, J.D., 2007. RNA regulons: coordination of post-transcriptional events. Nature Reviews Genetics 8, 533–543. doi:10.1038/nrg2111

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., Zavolan, M., 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat. Methods 8, 559–564. doi:10.1038/nmeth.1608

Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S., Green, R., 2015. Ribosomes slide on lysine-encoding homopolymeric A stretches. Elife 4, 446. doi:10.7554/eLife.05534

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., Ule, J., 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. 17, 909–915. doi:10.1038/nsmb.1838

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., Rajewsky, N., 2011. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol. Cell 43, 340–352. doi:10.1016/j.molcel.2011.06.008

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

Maatz, H., Jens, M., Liss, M., Schafer, S., Heinig, M., Kirchner, M., Adami, E., Rintisch, C., Dauksaite, V., Radke, M.H., Selbach, M., Barton, P.J.R., Cook, S.A., Rajewsky, N., Gotthardt, M., Landthaler, M., Hubner, N., 2014. RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. J. Clin. Invest. 124, 3419–3430. doi:10.1172/JCI74523

Mansfield, K.D., Keene, J.D., 2012. Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res. 40, 2734–2746. doi:10.1093/nar/gkr1114

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770. doi:10.1093/bioinformatics/btr011

Martin, G., Gruber, A.R., Keller, W., Zavolan, M., 2012. Genome-wide analysis of pre-mRNA 3“ end processing reveals a decisive role of human cleavage factor I in the regulation of 3” UTR length. Cell Rep 1, 753–763. doi:10.1016/j.celrep.2012.05.003

Mesarovic, M.D., Sreenath, S.N., Keene, J.D., 2004. Search for organising principles: understanding in systems biology. Syst Biol (Stevenage) 1, 19–27.

Moore, M.J., 2005. From birth to death: the complex lives of eukaryotic mRNAs. Science 309, 1514–1518. doi:10.1126/science.1111443

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., Ohler, U., 2017.

Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. 24, 86–96. doi:10.1038/nsmb.3325

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U., Keene, J.D., 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol. Cell 43, 327–339. doi:10.1016/j.molcel.2011.06.007

Mukherjee, N., Jacobs, N.C., Hafner, M., Kennington, E.A., Nusbaum, J.D., Tuschl, T., Blackshear, P.J., Ohler, U., 2014. Global target mRNA specification and regulation by the RNA-binding protein ZFP36. Genome Biol. 15, R12. doi:10.1186/gb-2014-15-1-r12

Munteanu, A., Mukherjee, N., Ohler, U., 2018. SSMART: Sequence-structure motif identification for RNA-binding proteins. bioRxiv 287953. doi:10.1101/287953

Nielsen, F.C., Nielsen, J., Christiansen, J., 2001. A family of IGF-II mRNA binding proteins (IMP) involved in RNA trafficking. Scand. J. Clin. Lab. Invest. Suppl. 234, 93–99.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., Fu, X.-D., 2013. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol. Cell 50, 223–235. doi:10.1016/j.molcel.2013.03.001

Pullmann, R., Kim, H.H., Abdelmohsen, K., Lal, A., Martindale, J.L., Yang, X., Gorospe, M., 2007. Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. Mol. Cell. Biol. 27, 6265–6278. doi:10.1128/MCB.00500-07

Sheth, U., Parker, R., 2003. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science 300, 805–808. doi:10.1126/science.1082320

Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X., Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., Smalec, B.M., Wang, X., Hong, E.L., Davidson, J.M., Lécuyer, E., Graveley, B.R., Yeo, G.W., 2016. Resources for the Comprehensive Discovery of Functional RNA Elements. Mol. Cell 61, 903–913. doi:10.1016/j.molcel.2016.02.012

Tenenbaum, S.A., Carson, C.C., Lager, P.J., Keene, J.D., 2000. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. U.S.A. 97, 14085–14090. doi:10.1073/pnas.97.26.14085

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., Darnell, R.B., 2003. CLIP identifies Nova-regulated RNA networks in the brain. Science 302, 1212–1215. doi:10.1126/science.1090095

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

Wilusz, J., Feig, D.I., Shenk, T., 1988. The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of polyadenylation cleavage sites. Mol. Cell. Biol. 8, 4477–4483.

Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J., Darnell, R.B., 2010. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science 329, 439–443. doi:10.1126/science.1191150

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E.,

790 Borowsky, M., Lee, J.T., 2010. Genome-wide identification of polycomb-associated RNAs
791 by RIP-seq. Mol. Cell 40, 939–953. doi:10.1016/j.molcel.2010.12.011
792