The architecture of the human RNA-binding protein regulatory network

Alessandro Quattrone and Erik Dassi*

Centre for Integrative Biology, University of Trento, Trento, Italy

*Correspondence to **erik.dassi@unitn.it**

Running head: *The RBP-RBP regulatory network*

Keywords: *RNA-binding proteins, post-transcriptional regulation, regulatory network, regulatory elements, cooperation, competition*

Abstract

RNA-binding proteins (RBPs) are key players of post-transcriptional regulation of gene expression. These proteins influence both cellular physiology and pathology by regulating processes ranging from splicing and polyadenylation to mRNA localization, stability, and translation. To fine-tune the outcome of their regulatory action, RBPs rely on an intricate web of competitive and cooperative interactions. Several studies have described individual interactions of RBPs with RBP mRNAs, suggestive of a RBP-RBP regulatory structure. Here we present the first systematic investigation of this structure, based on a network including almost fifty thousand experimentally determined interactions between RBPs and bound RBP mRNAs.

Our analysis identified two features defining the structure of the RBP-RBP regulatory network. What we call “RBP clusters” are groups of densely interconnected RBPs which co-regulate their targets, suggesting a tight control of cooperative and competitive behaviors. “RBP chains”, instead, are hierarchical structures driven by evolutionarily ancient RBPs, which connect the RBP clusters and could in this way provide the flexibility to coordinate the tuning of a broad set of biological processes.

The combination of these two features suggests that RBP chains may use the modulation of their RBP targets to coordinately control the different cell programs controlled by the RBP clusters. Under this *island-hopping* model, the regulatory signal flowing through the chains hops from one RBP cluster to another, implementing elaborate regulatory plans to impact cellular phenotypes. This work thus establishes RBP-RBP interactions as a backbone driving post-transcriptional regulation of gene expression to allow the fine-grained control of RBPs and their targets.

Introduction

In the last years post-transcriptional regulation of gene expression (PTR) has gained recognition as a crucial determinant of protein levels, and consequent cell phenotypes (Schwanhäusser et al. 2011; Vogel et al. 2010), stimulating a rising interest in studies focused on RNA-binding proteins (RBPs) and the interactions with their RNA targets.

RBPs are a key class of regulators in PTR. They are less than two thousand proteins in the human genome (almost 1200 verified RBPs plus several recently discovered ones (Castello et al. 2012)) and are made of modular domains of which RRM is the most represented one, found in over 200 proteins (Lunde et al. 2007). RBPs control processes ranging from splicing and polyadenylation to mRNA localization, stability, and translation (Gerstberger et al. 2014). To fine-tune the outcome of their regulatory action, RBPs rely on an intricate web of competitive and cooperative interactions (Dassi 2017).

Techniques such as ribonucleoprotein immunoprecipitation (RIP) and cross-linking and immunoprecipitation (CLIP) variants (Lee and Ule 2018) now allow us to identify the RNA targets of an RBP at the genome-wide scale. RBPs are involved in multiple aspects of physiology (e.g. brain and ovary development, immune response and the circadian cycle (Gerstberger et al. 2014; Lim and Allada 2013)) and pathology, being their alteration associated with a variety of diseases such as cancer, neurological and neuromuscular disorders (Wurth and Gebauer 2015; Lukong et al. 2008). The importance of obtaining a proper understanding of RBP properties and functions is thus evident.

While identifying the mRNA targets of RBPs, several works have highlighted among them an enrichment of mRNAs coding for gene expression regulators, including other RBPs but also transcription factors (TFs). This finding brought to the *regulator-of-regulators* concept (Keene 2007; Mansfield and Keene 2009), hinting at the existence of an extensive regulatory hierarchy of RBPs. For instance, we and others have specifically studied the *HuR*/*ELAVL1* protein (Dassi

et al. 2013; Mukherjee et al. 2011; Pullmann et al. 2007), which resulted to regulate the mRNAs of many RBPs (Mukherjee et al. 2011), several of which contain its same RNA-binding domain, the RRM. The increasing number of high-throughput data available is now allowing us to probe if this phenomenon occurs on a genome-wide scale. We chose to address this issue by specifically extracting the binding map of RBPs to their cognate mRNA and mRNAs of other RBPs. A similar approach has been previously applied for TF targets and metabolic networks in lower organisms such as *E. coli* and *S. cerevisiae* (Qijun Liu et al. 2009; Jothi et al. 2009; Pham et al. 2007); the human TF-TF regulatory interaction network, testing the *regulator-of-regulators* concept in TFs, has also been described for 41 cell types (Neph et al. 2012).

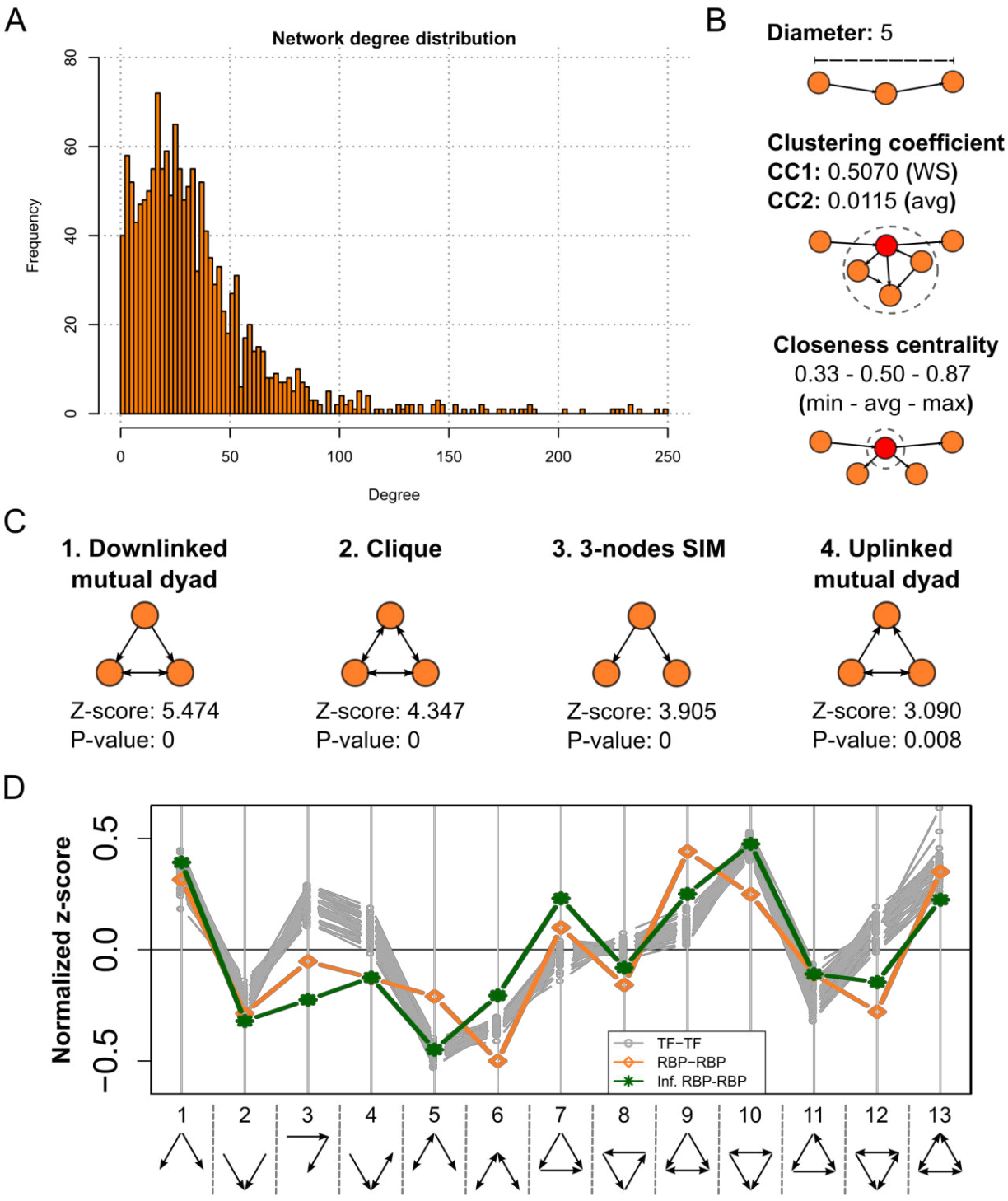
We present here the first systematic characterization of the RBP regulatory network, built by integrating experimental data on RBP targets. While sharing several properties of gene regulatory networks, its distinctive local structure hints at the specific dynamics of post-transcriptional regulation. We identified two major components which define the network structure. First, we found groups of densely connected RBPs which control each other to likely regulate cooperative and competitive behaviors on mutual targets. Then, we identified hierarchical node chains as the second feature shaping the network. In combination with RBP groups, these widespread regulatory units concur to the formation of a post-transcriptional backbone acting on multiple processes at once and could serve to coordinate major cell programs shaping cell phenotypes.

Results

Building the RBP-RBP network

Large-scale mapping of interactions between RBPs and their cognate mRNAs has been conducted by CLIP-like approaches (Lee and Ule 2018) in a few cellular systems, primarily HEK293, HeLa, and MCF7 cell lines. We previously collected these and other interactions in the

AURA 2 database (Dassi et al. 2014). We have now built the human RBP-RBP mRNA interaction network by extracting all related data and filtering each interaction by the expression of both interactors in the HEK293 cell line. To verify the generality of properties identified in this cell line, we have also constructed the same network for the other two cell lines with sufficient CLIP-like data, HeLa and MCF7. In our network vertices represent RBPs, and the presence of an edge between a source (protein) and a target (mRNA) RBP implies binding of the target RBP mRNA by the source RBP (which could result in post-transcriptional control of gene expression). The network includes 1536 RBPs out of 1827 (see methods for details on how we built the RBP list) connected by 47957 interactions. A total of 176 RBPs (11,5%) have outgoing interactions in the network (i.e., they bind the mRNA of an RBP) mostly coming from CLIP-like assays; the median network degree (number of connections) is 29, while the median number of individual binding sites for each RBP on each target RBP mRNAs is equal to 4. Among RBPs with outgoing interactions, 63 (35.8%) have self-loops (i.e., they bind their mRNA), confirming the general propensity of RBPs for autologous regulation. All interactions are listed in **Table S1**, and an interactive browser allowing to explore this and other networks is available at the AURA 2 (Dassi et al. 2014) website (http://aura.science.unitn.it).



111

112 **Figure 1: The RBP-RBP network is a gene regulatory network with a distinctive structure.**

113 **A)** shows the network degree distribution (up to 250), which follows a power-law. **B)** shows the

114 network diameter (top), its average clustering coefficients (middle, Watts-Strogatz 1-neighbor

coefficient, named CC1, and 2-neighbor coefficient, named CC2) and closeness centrality (bottom, minimum, average and maximum values for all nodes). **C)** shows the four most significant 3-nodes motifs identified by FANMOD with their z-score and p-value. **D)** displays the triad significance profile for the RBP-RBP network (orange line), the inferred RBP-RBP network (green line) and 41 TF-TF networks (gray lines). Positive z-scores indicate enrichment, negative ones depletion. While most motifs have similar z-scores in both networks, motifs 3, 4, 5, 9, 10 and 12 are differentially enriched in the RBP-RBP network, suggesting a distinctive structure with respect to the TF-TF networks.

The RBP-RBP network is a navigable "small-world" network

We first sought to verify whether the RBP-RBP network is a typical gene regulatory network, i.e. is "scale-free" and "small-world". To this end, we computed several global properties of the HEK293 cell network, shown in **Figure 1**. The degree distribution (**1A**) is following a power-law, with most nodes having a degree lower than 50 and a minor fraction reaching degrees over 200. This suggests that the network is scale-free, composed of a few central hubs and many progressively more peripheral nodes. The diameter (**1B,** D=5) indicates the network to be largely explorable by a few steps. Clustering coefficients (**1B**) suggest the presence of local-scale clustering (1-neighbor coefficient, CC1=0.507) which is lost when extending to more distant nodes (2-neighbor coefficient, CC2=0.0115). Eventually, closeness centrality (**1B**, C_c=0.5033) reiterates that most nodes are reachable by a small number of steps. We thus quantified this intuitive idea of network small-worldness by computing the S^{WS} measure (Humphries and Gurney 2008), which classifies a network as small-world when greater than 1. We obtained a value of 31.03, clearly supporting the hypothesis. Taken together, these values indeed put the network into the "small-world" class. Given its small diameter and high connectedness, the network can be considered navigable (Kleinberg 2000), i.e. apt to promote efficient information transmission along its paths. Eventually, we investigated the network

control structure (how it can be driven to any of its possible states), as described in (Ruths and Ruths 2014). We computed the network control profile, which resulted being [s=0.00367, e=0.99632, i=0.0], with *s* representing sources, *e* the external dilations and *i* the internal dilations. Hence, the network is dominated by external dilations (*e*), a fact that locates it in the class of top-down organization systems, structured to produce a correlated behavior throughout the system: members of this class are transcriptional networks, peer-to-peer systems, and corporate organizations (Ruths and Ruths 2014). These properties also hold in the HeLa and MCF7 networks, suggesting the stability of the network structure with different subsets of expressed RBPs (**Table S2**). We thus focused on the HEK293 network only for subsequent analyses.

RBP-RBP interactions define a hierarchical network structure

We then analyzed the local network structure by identifying motifs, i.e. recurrent patterns of RBP interaction. We used FANMOD (Wernicke and Rasche 2006) to look for 3-nodes motifs, of which several patterns have previously been characterized (e.g., the feed-forward loop and others (Milo et al. 2002)). The most significant motifs are shown in **Figure 1C**: among these, the *down-linked mutual dyad* (DMD) is the most enriched motif in our network. Together with the *single-input module* (SIM, third most enriched motif), these motifs indicate widespread use of hub-like patterns. The enrichment in DMD and *up-linked mutual dyad* (UMD, fourth most enriched motif), suggest a structure of ranked clusters for our network. Under this model, the dyads connect different hierarchical ranks within a network, with individual ranks structured as node clusters (1985; de Nooy et al. 2005). Instances of these motifs include *FXR2*, *HNRNPF*, and *TNRC6B* for the DMD, and *IGF2BP1*, *YWHAE,* and *YWHAG* for the SIM (with the first binding to the other two mRNAs). One realization of the UMD is that of *ELAVL1*, *LIN28B*, and *SYNCRIP* (e.g., both binding to *SYNCRIP* mRNA and the mRNA of each other).

We then identified 4-node motifs, the most significant of which are shown in **Supplementary Figure S1**. Among these, the forwarded uplinked mutual dyad is used to forward the output of an uplinked mutual dyad to a further RBP, and thus is a hierarchical, rank-connecting extension of the UMD. Furthermore, the chain-feeding dyad is made of a dyad which transmits its regulatory signal to two linearly connected RBPs, thus realizing a hierarchical structure as well. Given their properties, these two motifs provide further support to a ranked clusters model for the structure of our network.

The structure of the RBP-RBP network is different from the TF-TF one

We thus sought to compare the motif structure of the RBP-RBP network with the one of another network of regulators, the TF-TF network described in (Neph et al. 2012) for 41 cell types. We thus computed the triad significance profile (TSP) for these networks as described in (Milo et al. 2002). The TSP quantifies the use of the various three-nodes motifs by the network under analysis, and thus recapitulates its local structure. To complement this analysis, we also asked ourselves whether the structure of our network could be considered representative of the unavailable “complete” RBP-RBP network. To answer this question we thus built an inferred RBP-RBP network by collecting experimentally determined RBP-bound regions as per a protein occupancy profiling assay in HEK293 cells (Baltz et al. 2012). We then matched these regions to the binding motifs of 193 human RBPs derived from the *in vitro* RNAcompete assay (Ray et al. 2013). We obtained a network of 108161 RBP-RBP interactions. This network, independently reconstructed from two experimental datasets, becomes a validation of the general structure we propose for the RBP-RBP network.

We eventually compared the TSP of the three networks. The results are shown in **Figure 1D**, and we observe two salient aspects. First, the RBP-RBP network and its inferred version have a very similar motif structure (Pearson correlation=0.838, p-value=3.47e-04), with limited magnitude differences only, suggesting that our network structure is reproducible and a

192 representative cross-section of the complete set of RBP-RBP interactions. Then, the TF-TF
193 structure is instead more distant (mean Pearson correlation=0.719 across the 41 networks).
194 Indeed, 5/13 motifs are differentially represented in the RBP-RBP network (enriched instead of
195 depleted or vice versa), and the DMD is preferred over the UMD (the opposite being true for the
196 TF-TF networks). This suggests specialization of network structures by RBP-RBP interactions
197 with respect to TF-TF ones.
198
199 **The stoichiometry of RBP complexes is not determined by RBP-RBP regulatory**
200 **interactions**
201 We then asked if some type of biological constraints could be behind the evolutionary shaping
202 of the specific geometry of the RBP-RBP network. One likely hypothesis is that many
203 constraints are produced by interacting RBPs being part of the same ribonucleoprotein complex.
204 To test this hypothesis, we overlapped the interactions in our network with the experimental
205 binary protein-protein interactions (PPIs) contained in STRING (Szklarczyk et al. 2017), IntAct
206 (Orchard et al. 2014), and BioPlex (Huttlin et al. 2017). The low amount of network interactions
207 found to be mirrored in PPIs (3.37% for STRING, 0.57% for IntAct, and 0.29% for BioPlex)
208 suggests that the network wiring is not made to assure the availability of RBPs for complex
209 assembly. As this analysis dealt with single interactions, we then turned to whole complexes, as
210 stoichiometric ones (i.e., requiring precise quantities of each of the components for proper
211 functioning) may instead rely on this mechanism. We employed data from CORUM (Ruepp et
212 al. 2010) and found 1818 interactions overlapping a complex, corresponding to only 3.79% of
213 the network. **Table S3** lists complexes with at least five interactions in the network involving
214 their subunits. A few complexes are highly represented, including the large Drosha complex
215 (95% of its subunits are in the network, connected by 88 interactions) and the spliceosome
216 (83% of its subunits and 732 interactions). This suggests that only for some notable exceptions

stoichiometry of protein complexes is possibly driving the establishment of interactions in the RBP-RBP network.

Communities do not globally define the structure of the RBP-RBP network

To obtain a more general understanding of RBP-RBP interactions, we thus asked ourselves whether the network had a modular structure, made of RBP communities aimed at regulating specific biological processes. We used SurpriseMe (Aldecoa and Marin 2013), a tool for the investigation of community structures implementing several algorithms. SurpriseMe is based on Surprise (S) maximization (Aldecoa and Marin 2013), which has been shown to outperform the classic Girvan-Newman modularity measure Q (Newman and Girvan 2004). We used the communities identified by the two best-scoring algorithms implemented in the tool, namely CPM (Palla et al. 2005) and RNSC (King et al. 2004) (S=13698 and 13353 resp.). These algorithms detected a poor degree of modularity in the network: as shown in **Figure 2A**, 89% of the communities are singletons (i.e., formed by a single RBP) and only 8 contain more than 20 RBPs (1 with CPM and 7 with RNSC). Furthermore, both algorithms identified a huge community comprising a substantial portion of the network, suggesting a limited presence of true clustered structures. In that respect, the TF-TF networks appear to be much more modular, with much fewer communities identified as singletons (avg. of 53 vs. 657 for the RBP-RBP network) and higher average community size (avg. of 5 vs. 2.08). Eventually, we explored the enrichment of biological functions in the communities but detected no clear association involving most members of any of these. CPM and SCluster-derived communities are listed in **Table S4A** and **S4B**. Globally, these results suggest that the conventional community definition does not fit well the RBP-RBP network, which may thus be structured differently.

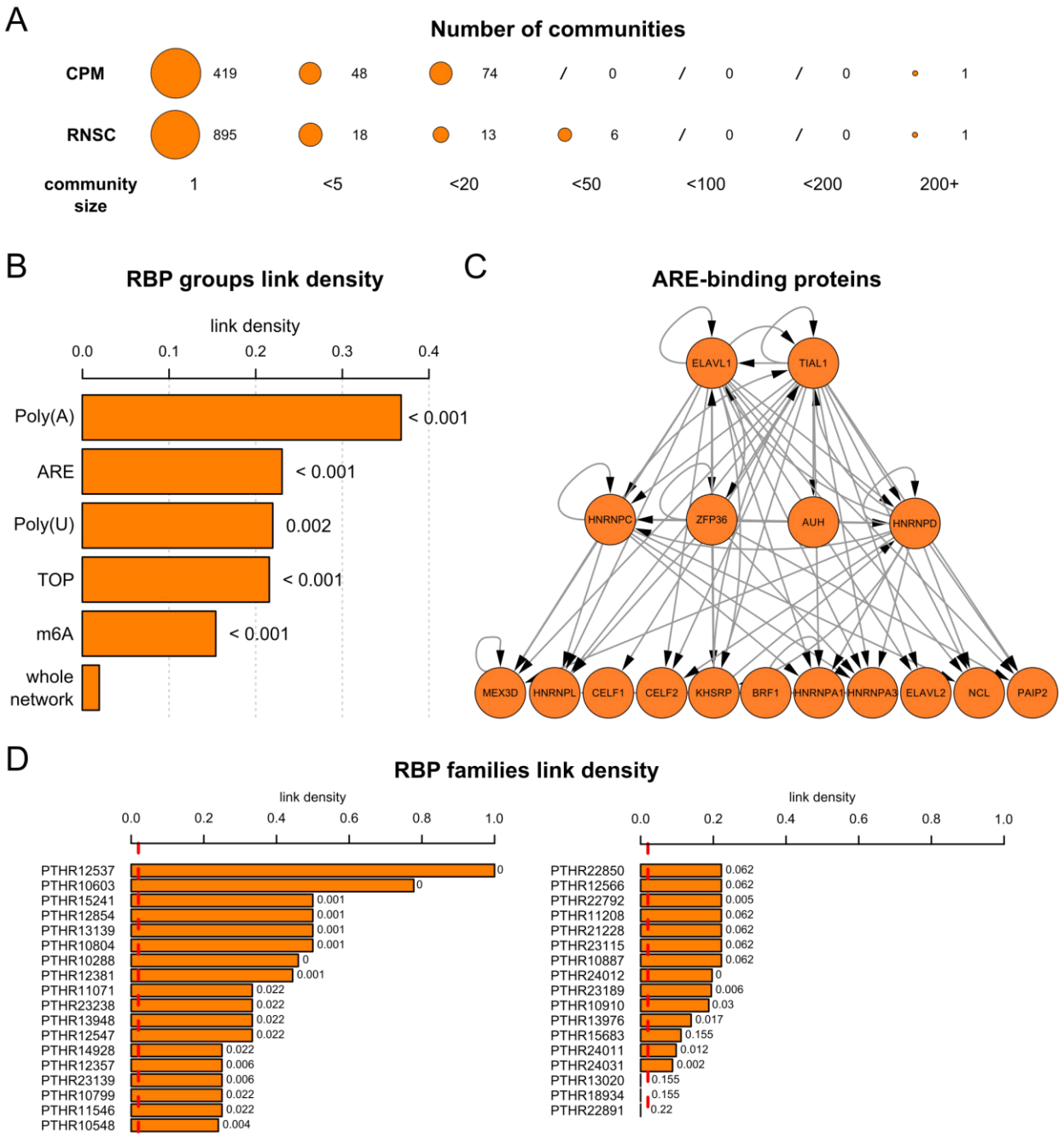


Figure 2: RBPs regulating common sets of targets are significantly more interconnected. **A)** displays the low modularity of the network as per the CPM and RNSC communities. Most are singletons, and one contains more than 25% of all RBPs. **B)** shows link density for the whole network and several groups of target-sharing RBPs: ARE, m6A, Poly(A), Poly(U) and TOP-binding proteins; 1000-samples bootstrap p-values are shown next to each bar. **C)** shows the

complete network of ARE-binding proteins, revealing a hierarchical structure dominated by *ELAVL1* and *TIAL1*. **D)** shows link density for families of RNA-binding proteins found in the network. A red dotted line indicates whole-network link density, and 1000-samples bootstrap p-values are shown next to each bar.

RBP-RBP interactions occur in clusters dictated by their common target mRNAs

The number and size of the detected communities indicate a low modularity of the RBP-RBP network, likely due to a peculiar community structure which cannot be detected by current algorithms. To further study this aspect, we set out to investigate a more general principle, that of interactions between RBPs in the network being connected to cooperatively or competitively sharing mRNA targets. RBP-RBP network wiring constraints could indeed be due to combinatorial RBP interactions through their targets (both RBPs, which are in the network, and non-RBPs, which are outside it). We thus extracted all mRNA targets for each RBP in the network from the AURA 2 database (Dassi et al. 2014) and computed the overlap for every RBP pair. We compared these overlaps for protein-mRNA pairs in the network (interacting RBPs) and pairs not in the network (non-interacting RBPs). The results indicated that interacting RBPs share significantly more targets than non-interacting RBPs (median 141 and 52 resp., Wilcoxon test $p<2.2E-16$). To investigate the biological meaning of this general phenomenon, we then studied sets of RBPs known to bind to the same cis-element and consequently sharing most, if not all, of their targets. We considered AU-Rich Element (ARE) binding proteins (Barreau et al. 2005), proteins interacting with the 5'UTR terminal oligopyrimidine tract (TOP) element (Tcherkezian et al. 2014; Hamilton et al. 2006), proteins interacting with the m6A RNA modification sites ((Roignant and Soller 2017)), and finally proteins interacting with with poly(U) RNAs, and with the poly(A), a major cis-determinant of mRNA stability and translation (Goss and Kleiman 2013). ARE-binding proteins, in particular, are known to display both cooperative and competitive behaviors (Barreau et al. 2005). We computed link density (i.e. the fraction of

all possible RBP-RBP interactions made within a group) for the whole network and each group. As shown in **Figure 2B**, all groups have significantly higher link densities than the whole network (7.8-18.7 times higher, 1000-samples bootstrap p-values=0.002 or less). The group with most interactions is the ARE-binding proteins (68 interactions), whose complete network is shown in **Figure 2C**. A hierarchical structure is visible, where *HuR*/*ELAVL1* and *TIAL1* are the major regulators (highest out-degree and lowest in-degree), connected to a second level (*ZFP36, HNRNPC*, *HNRNPD*, and *AUH),* which then controls the remaining RBPs (lowest out-degree).

Expanding on this idea, we eventually analyzed the link density of all annotated RBP families (as defined by Ensembl (Zerbino et al. 2018), see Methods). We assumed that, most often, members of the same RBP family cooperate or compete to regulate their common mRNA targets (Dassi 2017). Of the 288 families, 35 have at least two members in the network (i.e., taking part in at least one interaction). The median link density of these families is 0.24, with 32/35 having a higher density than the whole network (of these, 25 are significant according to a 1000-samples bootstrap) (**Figure 2D**). Although including only a fraction of all families, this results further indicate that RBP-RBP interactions may be needed to regulate cooperative and competitive behaviors on mutual targets, and that this behavior could be more prevalent than is currently known. "*RBP clusters*" (including families and sets of RBPs binding to the same cis-element) thus represent the community structure of the RBP-RBP network.

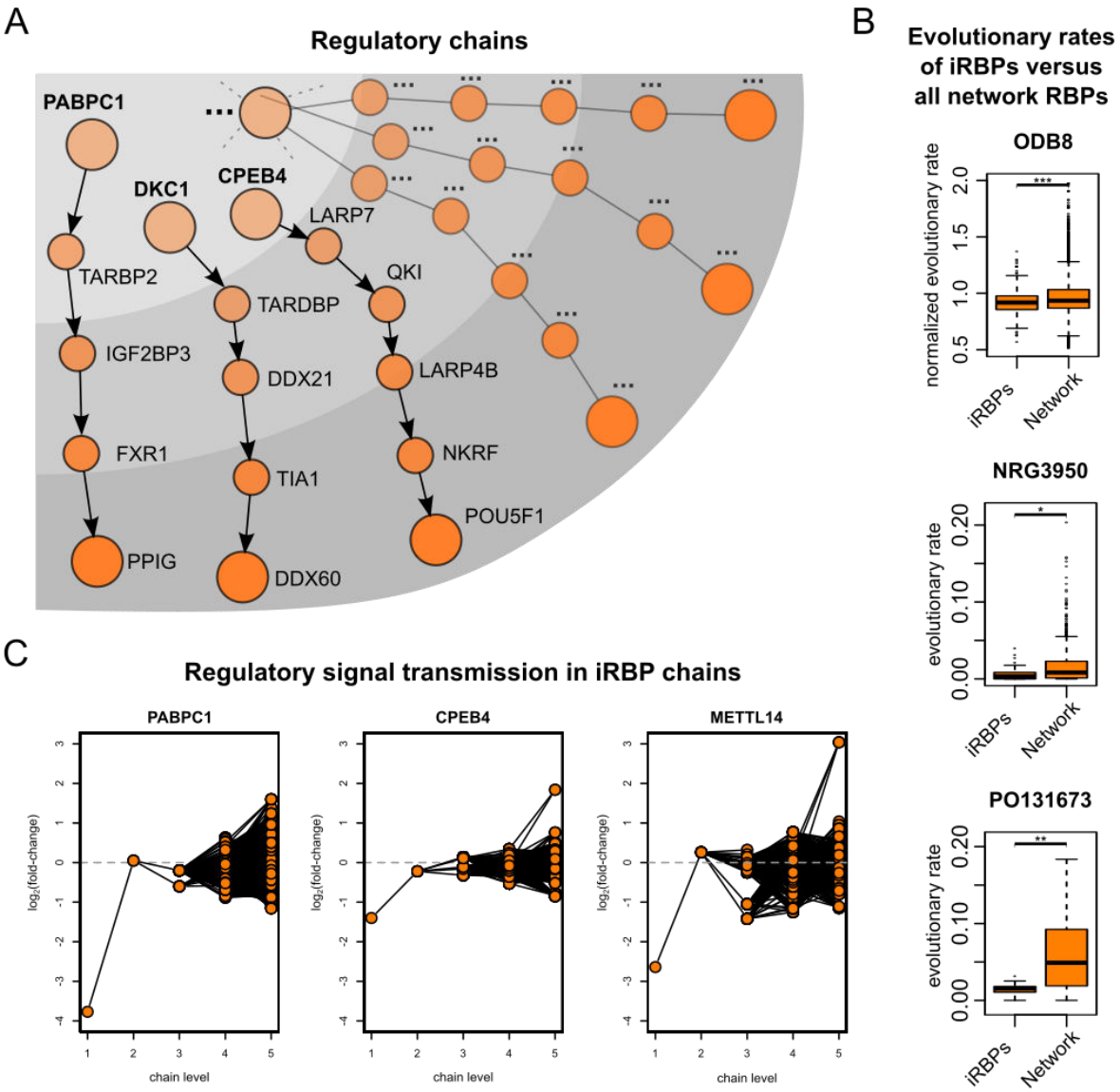


Figure 3: RBP chains dispatch regulatory information in the network. A) examples of RBP chains. Dashed lines and dotted name represent an iRBP heading many RBP chains. Increasing node color intensity represents the transmission of regulatory input through the chain, from the first to the last node**. B)** evolutionary rates of iRBPs and all RBPs in the network, obtained from the ODB8 database and two articles. iRBPs have a significantly lower rate in all datasets (Wilcoxon test p=5.2E-05, 0.0128 and 0.0016 for ODB8, NRG3950, and PO131673). **C)** Displays the log2 fold-change for RBPs at the various levels of chains led by PABPC1, CPEB4, and METTL14 when silencing these iRBPs. The first level of the chain is the silenced

iRBPs, while level 5 represents the last RBP of a chain (with levels 2..4 being intermediate steps of each chain). Lines represent the RBP-RBP connections in a chain while orange circles represent RBPs.

RBP chains are master regulatory units of the cell

The interactions identified by analyzing RBP clusters are, however, only a fraction of all links in the network. We thus hypothesized that, alongside these community-like structures, the network could also be employing linear node chains as its functional units. To study this aspect, we extracted chains of length 4 and 5 (longest network path) from the network (examples are shown in **Figure 3A**). To assess their relevance, we checked whether chains were more functionally homogeneous (i.e. composed of RBPs with more closely related functions) than algorithm-derived communities, taken as comparison given their poor ability to capture the structure of the RBP-RBP network. We thus computed a *functional coherence* score as the average semantic similarity score for each pair of RBPs in a chain or community. Chains display a significantly higher functional coherence than algorithm-derived communities (Wilcoxon test p=9.01E-07/0.0347 for CPM/RNSC for chains of length 4; p=7.562E-06/0.086 for CPM/RNSC for length 5; shown as density in **Figure S2**). Chains thus seem to be relevant to the RBP-RBP network organization. In the TF-TF network, chains are instead significantly less coherent (average 0.41/0.43 vs. 0.75/0.73 for TF-TF and RBP-RBP of length 4 and 5 resp.; Wilcoxon test p<2.2E-16 for both lengths), suggesting a lesser importance of such units in this network.

Chains are headed by a few initiator RBPs (iRBPs, 53 genes): these could be the most influential regulators, able to control many other RBPs and processes to dictate cell phenotypes. Therefore, iRBPs could be essential for the proper execution of cell processes. On this assumption, we searched for iRBPs in essential genes (defined by the underrepresentation of gene-trap vectors integration in their locus) of two human cell lines, as per a recent work (Blomen et al. 2015). As shown in **Table S1**, a third of the iRBPs are essential in both cell lines,

with 21 (40%) essential in the HAP1 line. iRBPs are enriched for essential genes in these cell models (max. Fisher test p=1.73E-05), and a 1000-samples bootstrap was significant (p<0.001) in both cell lines and their intersection. Most iRBPs (43/53, 81%) are also essential in at least one cellular model as per RNAi screenings included in the GenomeRNAi database (Schmidt et al. 2013). Merging all these annotations yields the remarkable total number of 46/53 iRBPs essential in at least one cell model (86%). To further strengthen this finding, we obtained the orthologs of iRBPs in mouse, D. *melanogaster,* and C. *elegans*, and compared them with essential genes in those organisms. As shown in **Table S5**, **S6**, and **S7**, the enrichment of essential genes in iRBPs is highly significant also for these organisms.

The iRBPs could also be highly conserved, due to their fundamental role in driving the RBP chains. We thus investigated whether these RBPs are more evolutionarily constrained than other RBPs. We extracted evolutionary rates of sequence divergence from the ODB8 database (Zdobnov et al. 2017) and (Zhang and Yang 2015), and rates of purifying selection from (Kryuchkova-Mostacci and Robinson-Rechavi 2015). We observed that, in all datasets, iRBPs have a significantly lower evolutionary rate than all RBPs in the network (**Fig 3B**; Wilcoxon test p=5.2E-05, 0.0128 and 0.0016 for ODB8, NRG3950 and PO131673 respectively). Furthermore, we also investigated whether the iRBPs PTR is evolutionarily constrained. We thus computed their average UTR conservation, and first found that UTRs of RBPs in the network are more conserved than the UTRs of all genes (Wilcoxon test p < 2.2E-16 for 5' and 3' UTRs). As the network includes most RBPs, this feature is characteristic of RBP genes. iRBPs 3'UTR conservation was also found to be significantly higher than that of other RBPs (Wilcoxon test p=0.002142). This ultra-conservation, coupled with the essentiality of most iRBPs, consistently support their importance as key cell regulators.

We eventually asked ourselves whether the regulatory information is transmitted through the chains, from the iRBPs down to the last node. To study this aspect, we obtained and reanalyzed publicly available transcriptome profiles of knock-down experiments for three iRBPs (two with

chains of length 4 and 5, *PABPC1* and *CPEB4*, and one with chains of length 4, *METTL14,* see Methods) in human cells. These RBPs act on various processes, including, for PABPC1 and METTL14, the regulation of mRNA stability. (Weng et al. 2018; Wang and Kiledjian 2000). We thus expect to detect at least a partial effect of their knock-down on these transcriptomes. We plotted the fold-change (knock-down versus control) of the RBPs composing the chains controlled by each of the three iRBPs. As shown in **Figure 3C**, a sizeable fraction of all chain members are differentially expressed (23.9% for PABPC1, 22.4% for CPEB4, and 46.3% for METTL14 at the adjusted p-value threshold of 0.05); if considering only a permissive fold-change threshold of 1.1 these numbers rise two to three times (66.1% for PABPC1, 44.1% for CPEB4, and 61.9% for METTL14). It must be noted that other modes of regulation, which cannot be observed in these datasets, can also be used by these proteins aside from mRNA stability (e.g., translational control). This data thus suggest that the regulatory information sparked by an iRBP is indeed transmitted through its chains, likely expanding the set of processes which can be controlled by these proteins. Chains are thus a functional unit in the RBP-RBP network, complementing the observed RBP clusters.

The RBP-RBP network is a robust and efficient hierarchy

We finally asked ourselves which were the implications of RBP chains on the global network structure. A reasonable hypothesis is that chains induce a hierarchical structure, as also suggested by the ranked clusters model we observed as defining the local network structure. We thus measured how hierarchical is the RBP-RBP network (Cheng et al. 2015), which revealed it as much more than any of the 41 TF-TF networks. When considering a hierarchy of 2, 4 or 6 levels; p-value is always orders of magnitude lower, with a $-\log_{10}p$ of 14.2 versus an average of 3.85 for TF-TF networks at six levels. Furthermore, feedback loops (not coherent with a hierarchical organization) are depleted in the network, representing 0.0085% of the motifs

379 only; feed-forward loops, coherent with a hierarchical organization, are instead enriched and
380 amount to 3.29% of the motifs.
381 We then assessed another desirable property, that of network robustness to the “removal” (i.e.,
382 loss of function) of an RBP from the network. To do so, we computed the pairwise
383 disconnectivity metrics on each node (Potapov et al. 2008). The metric is low (only 0.14% of
384 pairs are disconnected on average when removing a node from the network) and significantly
385 lower than for the TF-TF networks (average is three times higher for TFs, worst p-value=5.6E-
386 104). The network is thus well-tolerant to losing a node (fewer nodes are disconnected when
387 removing a node), which implies that RBP-RBP interactions are robust. This feature is likely
388 granted by the use of densely connected RBP clusters, resulting in partially redundant
389 regulation.
390 Eventually, while RBP clusters are redundant by definition (as they co-regulate a largely
391 overlapping set of targets), we asked whether also single RBP chains shared this property. We
392 thus computed the overlap between all targets (both RBPs, which are in the network, and non-
393 RBPs, which are outside it) of RBPs at the various levels of each chain of length 5. It resulted
394 being particularly low, as only 7.6% of the targets are overlapping between any two levels
395 (median of all chains, average of each pair in a chain; the range is 2.8%-15.5%). Differently
396 from RBP clusters, we can thus say that chains are efficient, as targets are not redundantly
397 regulated by individual RBPs along the chain, but rather are predominantly organized in
398 complementary sets at each of its nodes. This efficiency comes at the expense of robustness
399 (i.e., if one level of the chain fails the regulatory signal would most often be lost), which is
400 instead a feature of RBP clusters. The resulting model, shown in **Figure 4**, couples hierarchical
401 structure, network robustness through RBP groups, and efficiency through RBP chains.
402

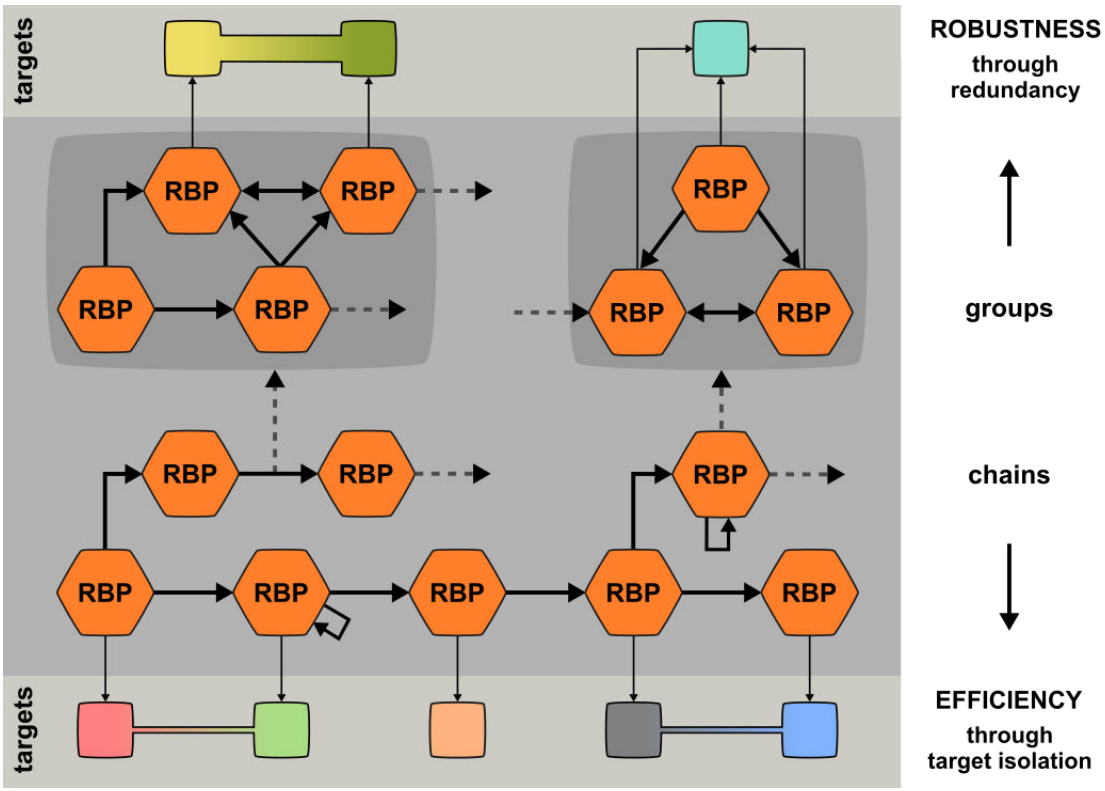


Figure 4: The RBP-RBP network is a robust and efficient hierarchy. Model of the RBP-RBP network as derived from our analyses. RBPs are indicated by hexagon-shaped nodes, RBP-RBP interactions by thick arrows (arrows pointing to the originating RBP represent autoregulation events), and targets sets by squares which can be shared by multiple RBPs (fraction of shared targets represented by the size of the shared area between the two squares). RBP-RBP interactions are robust due to densely connected RBP groups (co-regulating most of their targets), while RBP chains confer hierarchy and efficiency to the network, as target mRNA sets for each RBP in a chain are completely or predominantly different ("isolated"). Dashed in- and out-going arrows hint to the presence of further interactions within and between RBP groups and chains in the network.

Discussion

We presented here the first characterization of the RBP-RBP regulatory network. Starting from several reports hinting to a post-transcriptional hierarchy of regulators (Potapov et al. 2008; Dassi et al. 2013; Mukherjee et al. 2011; Pullmann et al. 2007), we collected available RBP-mRNA association data and described the network of interactions involving an RBP and an RBP mRNA. The network is small-world and scale-free, typical properties of gene regulatory networks. While the network is partial (as data is available for a fraction of all RBPs only), its local motif structure is highly coherent with the one of the inferred network, suggesting that it is representative of general patterns in RBP-RBP interactions.

Its local structure is similar to the one of TF-TF networks derived by DNase footprinting (Neph et al. 2012). However, differential enrichment of several motifs suggests that structure specialization occurs in the RBP-RBP network with respect to the TF-TF one, likely aimed at better suiting the specificities of post-transcriptional regulation. In particular, we found up- and down-linked mutual dyads as particularly enriched. These motifs are distinctive of a ranked clusters structure (Johnsen 1985), thus suggesting that the network can be divided in features conferring hierarchy and clusters of densely interacting nodes.

To study the role of these interactions in shaping cell phenotypes, we investigated why RBPs regulate each other. We found a few protein complexes involved in RNA metabolism and highly intra-regulated by RBP-RBP interactions. However, only a fraction of all complexes display this behavior, which cannot thus be considered general. We instead observed that groups of RBPs having overlapping targets tend to regulate each other: these interactions could represent a novel layer of regulation for cooperative and competitive behaviors between RBPs. As known for ARE-binding proteins (Barreau et al. 2005), RBPs can tune the expression of a common target by competitive or cooperative binding. We suggest that RBPs may influence the outcome of this process also by regulating the expression of the partner RBP, a mechanism which could

be used to reach precise ratios in a mRNP and yield the intended regulatory effect on common targets.

These groups represent “islands” of densely connected RBPs and are key in providing robustness to the network. Indeed, their partially redundant regulation improves the resilience of the network to the loss of function of individual RBPs.

However, RBP groups are not the only constituent feature of this network. Studying how RBPs interact with each other, we uncovered a set of widespread linear structures which appear to be more prevalent than communities. These structures, which we termed RBP chains and are driven by a few initiator RBPs (iRBPs), could provide enhanced flexibility with respect to a community pattern. Indeed, we believe that RBPs evolved the ability to influence a broad set of biological processes through such chains. Most iRBPs are essential for the cell, their 3’UTRs are more conserved, and their evolutionary rates are lower than for other RBPs. Taken together, these findings truly back iRBP importance as ancient master regulators of cell processes.

Chains profoundly shape the RBP-RBP network to be highly hierarchical. Their regulatory action confers efficiency, as the fraction of targets shared between the different chain levels is limited (i.e., regulation is not replicated along a chain), thus streamlining the flow of regulatory information from iRBPs to the final chain targets.

We have thus identified the two features hypothesized by the ranked clusters model: a hierarchy-inducing structure (the RBP chains), and clusters of densely interacting nodes (RBP groups). This indicates that this model fits well with the RBP-RBP network and can be found at different depths of observation: from the local, three-nodes motif structure, to the patterns defining the topology of the global network. The combination of properties offered by these features, namely robustness and efficiency, reflects the constant evolutionary pressure shaping a machinery as fundamental to the cell as is the one driving post-transcriptional regulation of gene expression. While establishing robustness only through RBP groups could in principle lead to a weaker architecture (as the regulatory signal going through chains is not replicated), it may

be cheaper to obtain and potentially more far-reaching. This consideration, however, raises a question: which is the role of chains in relation to RBP groups?

We suggest that RBP chains use the modulation of RBP targets as a connector to different processes, represented by the RBP groups. We call this model "*island-hopping*". Under this model, the regulatory signal originated from the chains iRBPs hops from one island (RBP group, side-connected to the chain) to another, while flowing through the chain levels, to regulate several cellular processes. This pattern thus allows potentially coordinating a broad set of functions of interest. Activating different chains would then result in the modulation of a different set of processes, granting substantial flexibility to the RBP-RBP network. This work thus establishes interactions among RBPs and RBP mRNAs as a backbone driving post-transcriptional regulation of gene expression to coordinately tune protein abundances.

Materials and Methods

RBP list construction and annotation

We built the list of human RNA-binding proteins by first extracting genes annotated as RNA-binding (GO:0003723) and protein-coding in Ensembl v92 (Zerbino et al. 2018), then merging these genes with the curated RBP list from a recent work (Sebestyén et al. 2016). The resulting list thus includes canonical and novel RBPs for a total of 1827 proteins. Families of RNA-binding proteins were extracted from Ensembl v92 gene families (Zerbino et al. 2018), by considering only families including more than one RBP.

Network construction

Regulatory interactions involving two RBPs were extracted from the AURA 2 database (Dassi et al. 2014). Interactions were filtered by requiring the expression of both participants in HEK293, HeLa or MCF7 cells, systems where the majority of the data were derived. Expressed genes

493 were determined by RNA-seq profiles of HEK293 (Kishore et al. 2011), HeLa (Cabili et al. 2011)
494 and MCF7 (Vanderkraats et al. 2013), using an expression threshold of 0.1 RPKM. The
495 direction of edges in the network represents regulation by the source RBP on the target RBP
496 mRNA.
497 The inferred RBP-RBP network was built by collecting RBP-bound regions in mRNA UTRs from
498 a protein occupancy profiling assay in HEK293 cells (Dassi et al. 2014; Baltz et al. 2012).
499 RNAcompete-derived PWMs for 193 human RBPs (Ray et al. 2013) were downloaded from
500 CISBP-RNA (Ray et al. 2013). Binding regions of these RBP on protein-bound regions were
501 identified by Biopython (Ray et al. 2013; Cock et al. 2009), selecting the best matching RBP for
502 each region (score threshold=0.99); only interactions involving two RBPs (one binding to the
503 other mRNA) were included in the network.
504 The networks were deposited in NDEX with ID *ee3e8898-6e29-11e8-a4bf-0ac135e8bacf*,
505 *f5ad750b-6e29-11e8-a4bf-0ac135e8bacf* and *fc1e526e-6e29-11e8-a4bf-0ac135e8bacf*.
506
507 **Network properties analysis**
508 Network diameter, degree distribution, closeness centrality, Watts-Strogatz (CC1) and two-
509 neighbor (CC2) clustering coefficient were computed by Pajek (Batagelj and Mrvar 2002) and
510 plotted with R (Tierney 2012). The S^{WS} measure was computed as described in (Humphries and
511 Gurney 2008) by using the Watts-Strogatz clustering coefficient and generating the required
512 random network with Pajek (Batagelj and Mrvar 2002). The network control structure was
513 computed by Zen (Ruths and Ruths 2014). Hierarchical score was computed as per (Cheng et
514 al. 2015) and pairwise disconnectivity obtained by DiVa (Potapov et al. 2008).
515 Link density for a set of nodes was computed as (number of links between nodes in the set) /
516 (number of nodes in the set^2). Bootstraps were performed by 1000 random selections of a
517 number of nodes equal to the set size and computation of the link density for each of these.
518

519 **Network structure analysis**

520 Network motifs of size 3 and 4 were identified with FANMOD (Cheng et al. 2015; Wernicke and
521 Rasche 2006) using 1000 random networks (100 for motifs of size 4, due to required computing
522 time), 3 exchanges per edge and 3 exchange attempts. Triad significance profiles for motifs of
523 size 3 were computed as described in (Milo et al. 2004) for the RBP-RBP network, the inferred
524 RBP-RBP network and the TF-TF networks described in (Neph et al. 2012).
525 Communities were studied with the SurpriseMe tool (Aldecoa and Marin 2013): CPM (Palla et
526 al. 2005) and RNSC (King et al. 2004), the algorithms obtaining the highest S values, were used
527 to define communities. Chains were extracted from the network with igraph (http://igraph.org);
528 functional coherence scores were computed with GOSimSem (Yu et al. 2010) by averaging the
529 semantic similarity of each pair of genes in the chain.
530
531 **Protein-protein interactions and complexes overlap**
532 Human protein-protein interactions were extracted from STRING (Szklarczyk et al. 2017),
533 BioPlex (Huttlin et al. 2017), and IntAct (Orchard et al. 2014), retaining only interactions of the
534 “binding” type (physical association) and with both partners being in our network. Human protein
535 complexes were downloaded from CORUM (Ruepp et al. 2010). Overlaps were performed by
536 custom Python scripts.
537
538 **Gene essentiality and phylogenetic conservation analysis**
539 Essential genes of human cells were obtained from (Blomen et al. 2015). Genes associated with
540 an embryonic lethal phenotype in mouse from the MGI (Blake et al. 2017); genes associated
541 with a lethality phenotype were extracted from WormBase (Lee et al. 2018) and FlyBase
542 (Gramates et al. 2017) for C.*elegans* and D.*melanogaster* respectively. Orthologs of iRBPs
543 were extracted from the same databases. Bootstraps were computed by 1000 random

544 selections of as many genes as iRBPs and computing the fractions of these in the essential
545 genes for each organism.
546 UTR conservation scores were computed by averaging the phastCons score derived from the
547 UCSC 46-way vertebrate alignment (Casper et al. 2018). The average score of all 5' or 3' UTRs
548 of a gene was employed as the conservation score for that gene 5' or 3'UTRs. Protein
549 evolutionary rates were obtained from the ODB8 database (Zdobnov et al. 2017) and two
550 articles (Kryuchkova-Mostacci and Robinson-Rechavi 2015; Zhang and Yang 2015); statistical
551 tests were performed by R (Tierney 2012).
552
553 **iRBP knock-down datasets**
554 RNA-seq datasets following the knock-down of PABPC1, CPEB4, and METTL14 were obtained
555 from GEO (IDs: GSE88099, GSE88545, and GSE56010). Reads were quality-trimmed and
556 adapters removed with Trimmomatic (Bolger et al. 2014), then aligned to the human genome
557 (hg38 assembly), and transcripts quantified (Gencode v28 annotation) with STAR (Dobin et al.
558 2013). Differential expression was eventually computed with DESeq2 (Love et al. 2014) using
559 an adjusted p-value threshold of 0.05.
560
561
562
563
564
565
566

Supplemental Figures

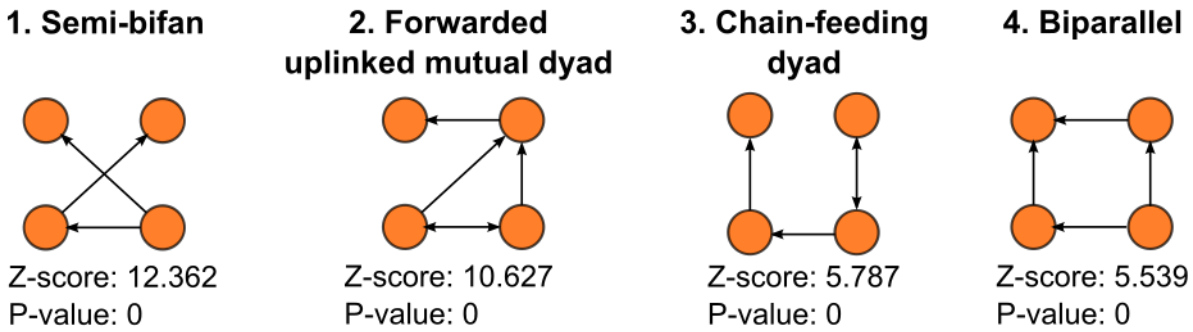


Figure S1: Most enriched 4-nodes motifs in the network. Shows the four most significant 4-nodes motifs identified by FANMOD with their z-score and p-value.

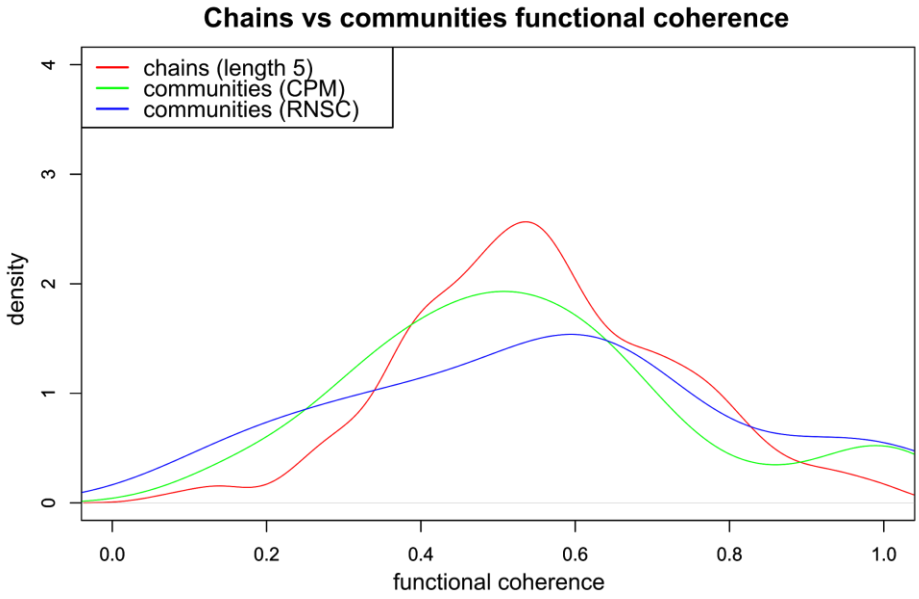


Figure S2: Functional coherence distribution for chains versus communities. Shows the differences in density of the functional coherence values distribution between chains of length 6 (in red) and communities obtained by CPM (green) and RNSC (blue).

Acknowledgments

-

Competing financial interests

The authors declare no competing financial interests.

References

Aldecoa R, Marin I. 2013. SurpriseMe: an integrated tool for network community structure characterization using Surprise maximization. Bioinformatics 30: 1041–1042.

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**: 674–690.

Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* **33**: 7138–7150.

Batagelj V, Mrvar A. 2002. Pajek— Analysis and Visualization of Large Networks. In *Lecture Notes in Computer Science*, pp. 477–478.

Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, the Mouse Genome Database Group. 2017. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Res* **45**: D723–D729.

Blomen VA, Májek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A, et al. 2015. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**: 1092–1096.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915–1927.

602 Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee
603 BT, Karolchik D, et al. 2018. The UCSC Genome Browser database: 2018 update. *Nucleic*
604 *Acids Res* **46**: D762–D769.
605 Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys
606 DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of
607 mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.
608 Cheng C, Andrews E, Yan K-K, Ung M, Wang D, Gerstein M. 2015. An approach for
609 determining and measuring network hierarchy applied to comparing the phosphorylome
610 and the regulome. *Genome Biol* **16**: 63.
611 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff
612 F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational
613 molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.
614 Dassi E. 2017. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins.
615 *Front Mol Biosci* **4**: 67.
616 Dassi E, Re A, Leo S, Tebaldi T, Pasini L, Peroni D, Quattrone A. 2014. AURA 2. *Translation* **2**:
617 e27738.
618 Dassi E, Zuccotti P, Leo S, Provenzani A, Assfalg M, D'Onofrio M, Riva P, Quattrone A. 2013.
619 Hyper conserved elements in vertebrate mRNA 3'-UTRs reveal a translational network of
620 RNA-binding proteins controlled by HuR. *Nucleic Acids Res* **41**: 3201–3216.
621 de Nooy W, Mrvar A, Batagelj V. 2005. *Exploratory Social Network Analysis with Pajek*.
622 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
623 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
624 Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. *Nat Rev*
625 *Genet* **15**: 829–845.
626 Goss DJ, Kleiman FE. 2013. Poly(A) binding proteins: are they all created equal? *Wiley*
627 *Interdiscip Rev RNA* **4**: 167–179.

628 Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ,
629 Tabone CJ, Crosby MA, Emmert DB, et al. 2017. FlyBase at 25: looking to the future.
630 *Nucleic Acids Res* **45**: D663–D671.
631 Hamilton TL, Stoneley M, Spriggs KA, Bushell M. 2006. TOPs and their regulation. *Biochem*
632 *Soc Trans* **34**: 12–16.
633 Humphries MD, Gurney K. 2008. Network "small-world-ness": a quantitative method for
634 determining canonical network equivalence. *PLoS One* **3**: e0002051.
635 Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP,
636 Parzen H, et al. 2017. Architecture of the human interactome defines protein communities
637 and disease networks. *Nature* **545**: 505–509.
638 Johnsen EC. 1985. Network macrostructure models for the Davis-Leinhardt set of empirical
639 sociomatrices. Soc Networks 7: 203–224.
640 Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Babu MM. 2009.
641 Genomic analysis reveals a tight link between transcription factor dynamics and regulatory
642 network architecture. *Mol Syst Biol* **5**: 294.
643 Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**:
644 533–543.
645 King AD, Przulj N, Jurisica I. 2004. Protein complex prediction via cost-based clustering.
646 *Bioinformatics* **20**: 3013–3020.
647 Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. 2011. A quantitative
648 analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*
649 **8**: 559–564.
650 Kleinberg JM. 2000. Navigation in a small world. *Nature* **406**: 845.
651 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2015. Tissue-Specific Evolution of Protein
652 Coding Genes in Human and Mouse. *PLoS One* **10**: e0131673.
653 Lee FCY, Ule J. 2018. Advances in CLIP Technologies for Studies of Protein-RNA Interactions.

Mol Cell **69**: 354–369.

Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Davis P, Gao S, Grove C, et al. 2018. WormBase 2017: molting into a new stage. *Nucleic Acids Res* **46**: D869–D874.

Lim C, Allada R. 2013. Emerging roles for post-transcriptional regulation in circadian clocks. *Nat Neurosci* **16**: 1544–1550.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Lukong KE, Chang K-W, Khandjian EW, Richard S. 2008. RNA-binding proteins in human genetic disease. *Trends Genet* **24**: 416–425.

Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**: 479–490.

Mansfield KD, Keene JD. 2009. The ribonome: a dominant force in co-ordinating gene expression. *Biol Cell* **101**: 169–181.

Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U. 2004. Superfamilies of evolved and designed networks. *Science* **303**: 1538–1542.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002. Network motifs: simple building blocks of complex networks. *Science* **298**: 824–827.

Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M Jr, Tuschl T, Ohler U, Keene JD. 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43**: 327–339.

Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. 2012. Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150**: 1274–1286.

Newman MEJ, Girvan M. 2004. Finding and evaluating community structure in networks.

Physical Review E **69**. http://dx.doi.org/10.1103/physreve.69.026113.

Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, et al. 2014. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**: D358–63.

Palla G, Derényi I, Farkas I, Vicsek T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**: 814–818.

Pham H, Ferrari R, Cokus SJ, Kurdistani SK, Pellegrini M. 2007. Modeling the regulatory network of histone acetylation in Saccharomyces cerevisiae. *Mol Syst Biol* **3**: 153.

Potapov AP, Goemann B, Wingender E. 2008. The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks. *BMC Bioinformatics* **9**: 227.

Pullmann R Jr, Kim HH, Abdelmohsen K, Lal A, Martindale JL, 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.

Qijun Liu, Liu Q, Wang Z, Li M, Li D, Zhu Y. 2009. Revealing the functional modularity of yeast transcriptional regulatory network by using a novel topological measurement. In *2009 2nd IEEE International Conference on Computer Science and Information Technology* http://dx.doi.org/10.1109/iccsit.2009.5234476.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172–177.

Roignant J-Y, Soller M. 2017. mA in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. *Trends Genet* **33**: 380–390.

Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes H-W. 2010. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res* **38**: D497–501.

Ruths J, Ruths D. 2014. Control profiles of complex networks. *Science* **343**: 1373–1376.

706 Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M. 2013. GenomeRNAi: a database
707 for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res* **41**: D1021–6.
708 Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011.
709 Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.
710 Sebestyén E, Singh B, Miñana B, Pagès A, Mateo F, Pujana MA, Valcárcel J, Eyras E. 2016.
711 Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils
712 novel cancer-relevant splicing networks. *Genome Res* **26**: 732–744.
713 Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT,
714 Roth A, Bork P, et al. 2017. The STRING database in 2017: quality-controlled protein-
715 protein association networks, made broadly accessible. *Nucleic Acids Res* **45**: D362–D368.
716 Tcherkezian J, Cargnello M, Romeo Y, Huttlin EL, Lavoie G, Gygi SP, Roux PP. 2014.
717 Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of
718 5'TOP mRNA translation. *Genes Dev* **28**: 357–371.
719 Tierney L. 2012. The R Statistical Computing Environment. In *Lecture Notes in Statistics*, pp.
720 435–447.
721 Vanderkraats ND, Hiken JF, Decker KF, Edwards JR. 2013. Discovering high-resolution
722 patterns of differential DNA methylation that correlate with gene expression changes.
723 *Nucleic Acids Res* **41**: 6816–6827.
724 Vogel C, Abreu R de S, Ko D, Le S-Y, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte
725 EM, Penalva LO. 2010. Sequence signatures and mRNA concentration can explain two-
726 thirds of protein abundance variation in a human cell line. *Mol Syst Biol* **6**: 400.
727 Wang Z, Kiledjian M. 2000. The poly(A)-binding protein and an mRNA stability protein jointly
728 regulate an endoribonuclease activity. *Mol Cell Biol* **20**: 6334–6341.
729 Weng H, Huang H, Wu H, Qin X, Zhao BS, Dong L, Shi H, Skibbe J, Shen C, Hu C, et al. 2018.
730 METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes
731 Leukemogenesis via mRNA mA Modification. *Cell Stem Cell* **22**: 191–205.e9.

732 Wernicke S, Rasche F. 2006. FANMOD: a tool for fast network motif detection. *Bioinformatics*
733 **22**: 1152–1153.
734 Wurth L, Gebauer F. 2015. RNA-binding proteins, multifaceted translational regulators in
735 cancer. *Biochim Biophys Acta* **1849**: 881–886.
736 Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. 2010. GOSemSim: an R package for measuring
737 semantic similarity among GO terms and gene products. *Bioinformatics* **26**: 976–978.
738 Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M,
739 Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional
740 annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids*
741 *Res* **45**: D744–D749.
742 Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A,
743 Girón CG, et al. 2018. Ensembl 2018. *Nucleic Acids Res* **46**: D754–D761.
744 Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev*
745 *Genet* **16**: 409–420.
746