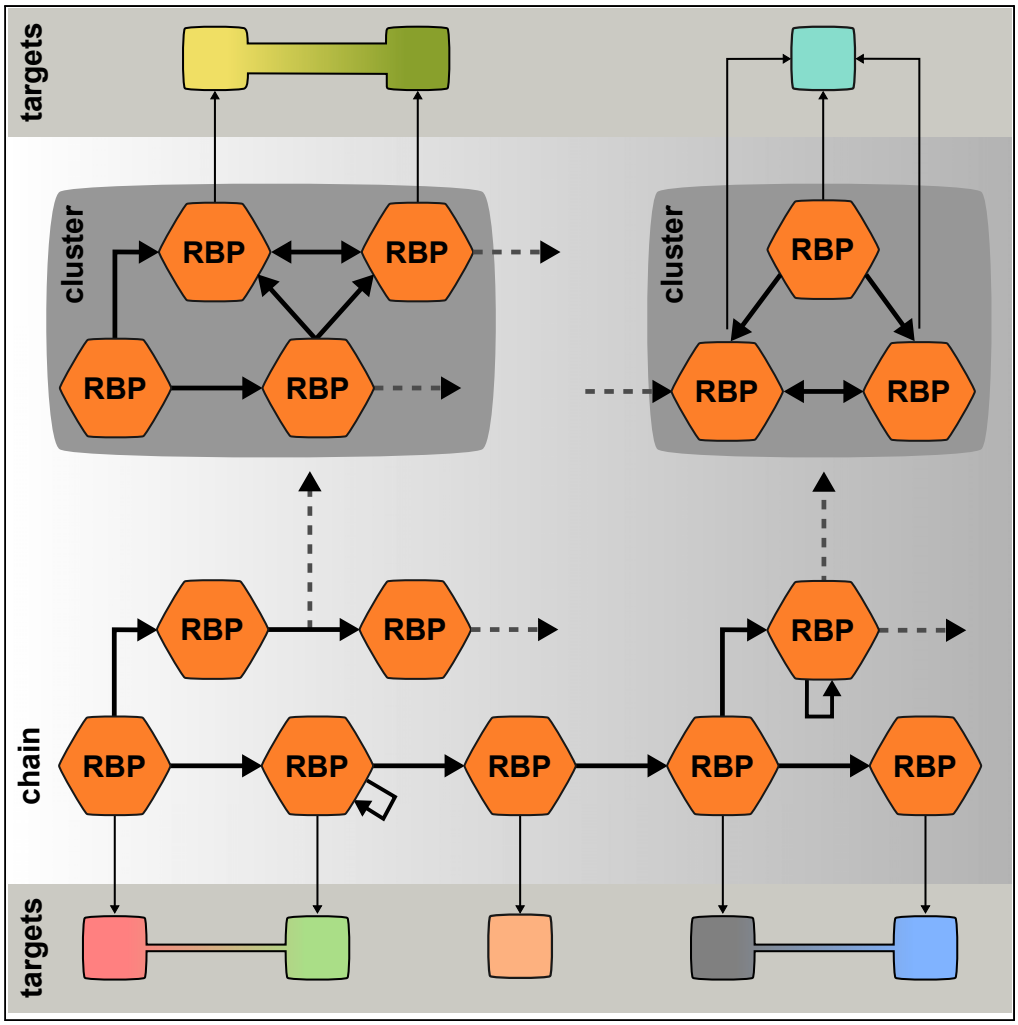
iScience

CellPress

Article

The Architecture of the Human RNA-Binding Protein Regulatory Network



Alessandro Quattrone, Erik Dassi

erik.dassi@unitn.it

HIGHLIGHTS

The RBP-RBP network is a robust and efficient hierarchical structure

The RBP-RBP network is formed by RBP clusters and chains

RBP-RBP interactions cluster to cooperate and compete on common target mRNAs

RBP chains are master regulatory units of the cell led by ancient RBPs

Quattrone & Dassi, iScience *21*, 706–719
November 22, 2019
https://doi.org/10.1016/j.isci.2019.10.058

iScience

CellPress

Article

The Architecture of the Human RNA-Binding Protein Regulatory Network

Alessandro Quattrone[1] and Erik Dassi[1,2,*]

SUMMARY

RNA-binding proteins (RBPs) are key players of post-transcriptional regulation of gene expression, relying on competitive and cooperative interactions to fine-tune their action. Several studies have described individual interactions of RBPs with RBP mRNAs. Here we present a systematic network investigation of fifty thousand interactions between RBPs and the UTRs of RBP mRNAs. We identified two structural features in this network. RBP clusters are groups of densely interconnected RBPs co-binding their targets, suggesting a tight control of cooperative and competitive behaviors. RBP chains are hierarchical structures connecting RBP clusters and driven by evolutionarily ancient RBPs. These features suggest that RBP chains may coordinate the different cell programs controlled by RBP clusters. Under this model, the regulatory signal flows through chains from one cluster to another, implementing elaborate regulatory plans. This work thus suggests RBP-RBP interactions as a backbone driving post-transcriptional regulation of gene expression to control RBPs action on 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). This has stimulated 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 2,000 proteins in the human genome (almost 1,200 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]). Furthermore, RBPs play a role in pathology. Their alteration is indeed associated with diseases such as cancer and 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 were found 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 datasets 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 *Escherichia coli* and *Saccharomyces cerevisiae* (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). The study of a cross-regulatory network

[1]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, TN 38123, Italy

[2]Lead Contact

*Correspondence: erik.dassi@unitn.it

https://doi.org/10.1016/j.isci.2019.10.058

of transcription factors, kinases, and splicing factors has been attempted by Kosti et al. (2012), narrower in scope as it included only 20 RBPs. Also, Mukherjee et al. (2019) have recently studied the properties of the RBP-target RNA network, although including only PAR-CLIP data and considering all interactions (not only RBP-RBP ones).

We present here a systematic characterization of the RBP-RBP regulatory network, built by integrating experimental data on RBPs targeting the UTRs of RBP mRNAs derived by multiple techniques. 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 that define the network structure. First, we found groups of densely connected RBPs that 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 clusters, these widespread regulatory units concur to the formation of a post-transcriptional backbone. This backbone acts on multiple processes at once and could serve to coordinate major cell programs.

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 interaction network of RBPs and RBPs mRNAs (thereafter the RBP-RBP network) by extracting all related data from AURA 2 and filtering each interaction by the expression of both interactors in the HEK293 cell line (see Methods). To verify the generality of properties identified in this cell line, we have also constructed the same network for two other frequently used cell lines, HeLa and MCF7. To do so, we filtered the interactions by the expression of both partners (the binding RBP and the target gene mRNA) in those two cell lines, thus retaining only possible interactions (i.e., interactions for which both partners are expressed in that cell line). In our network, vertices represent RBPs, and edges are post-transcriptional interactions. 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 1,536 RBPs out of 1,827 (see Methods for details on how we built the RBP list) connected by 47,957 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, whereas 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. An example of this behavior is that of *CPEB4*, previously described to repress its own mRNA to control terminal erythropoiesis (Hu et al., 2014). All interactions are listed in Table S1. 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), and the networks were also deposited in NDEX (http://www.ndexbio.org, see Methods).

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., "scale-free" and "small-world." To this end, we computed several global properties of the HEK293 network (Figure 1). The degree distribution (1A) follows 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. If considering the out- and in-degree separately, the former appears strongly scale-free, whereas the latter is weaker in this respect, as would also be expected for transcription factors networks. 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). Furthermore, 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). By comparing the network with its randomly wired counterpart, this measure 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, as is most often the case for the broad class of gene regulatory networks. Given its small diameter and high connectedness, the network can be considered navigable

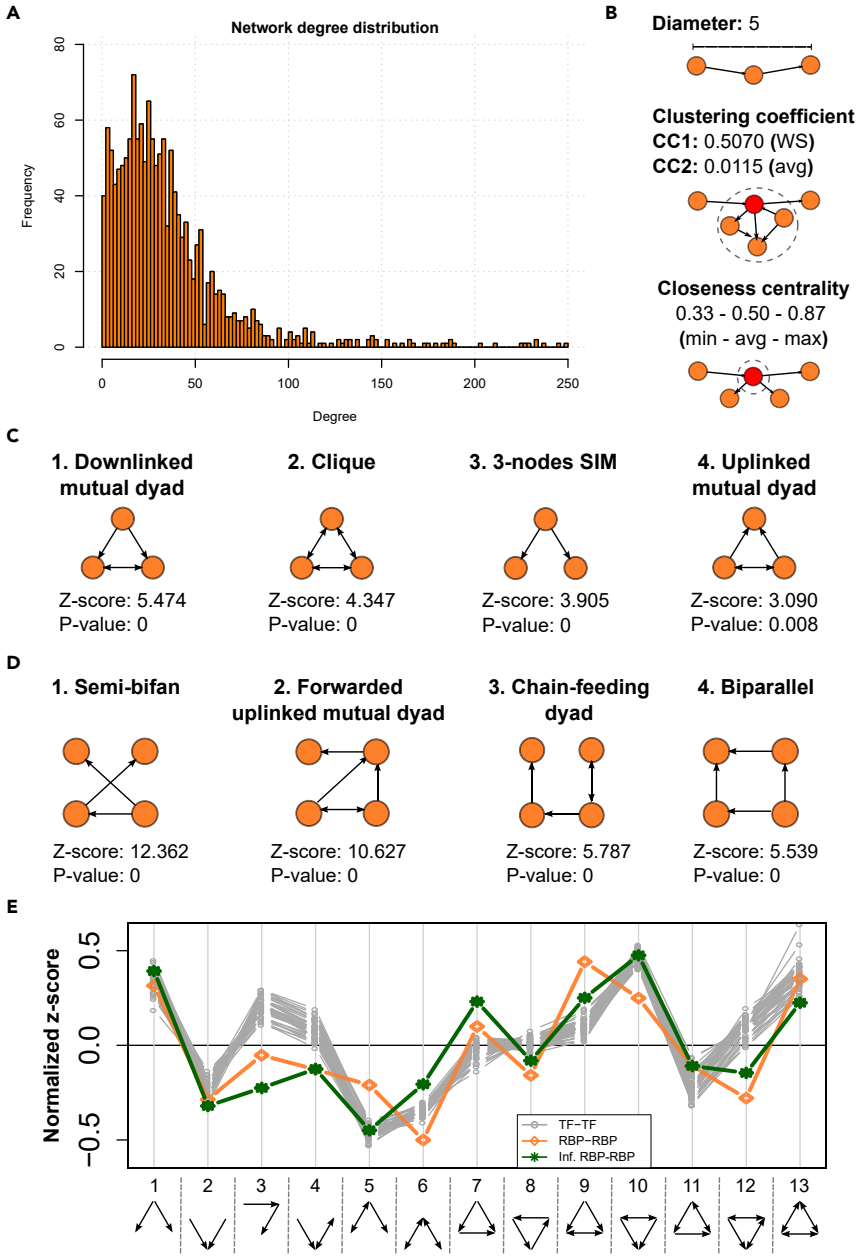


Figure 1. The RBP-RBP Network Is a Gene Regulatory Network with a Distinctive Structure

(A) Network degree distribution (up to 250), following a power-law distribution.

(B) The 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) The four most significant three-node motifs identified by FANMOD with their *Z* score and p value.

(D) The four most significant four-node motifs identified by FANMOD with their *Z* score and p value.

(E) 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 z-scores depletion. Although 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.

See also Tables S1 and S2.

(Kleinberg, 2000), i.e., apt to promote efficient information transmission along its paths. Eventually, we investigated the network control structure (how information can flow across it, as described in Ruths and Ruths, 2014). We computed the network control profile, which was found to be [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. Such networks induce a correlated behavior throughout the system: members of this class are transcriptional networks, peer-to-peer systems, and corporate organizations (Ruths and Ruths, 2014). The TF-TF networks have a similar profile (0.261, 0.699, and 0.04 for the Renal Cortical Epithelial Cell one), although less imbalanced on external dilations. 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 interaction between RBPs and RBPs mRNAs. We used FANMOD (Wernicke and Rasche, 2006) to look for three-node motifs, of which several patterns have previously been characterized (e.g., the feedforward 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 (de Nooy et al., 2005; Johnsen, 1985). 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 mRNA of the other two). One example of the UMD is that of *ELAVL1*, *TIAL1*, and *TIA1* (e.g., both binding to *TIA1* mRNA and each other mRNA), whose interactions were previously described to be functional, with *ELAVL1* promoting *TIA1* expression while *TIAL1* exerts the opposite effect (Pullmann et al., 2007).

We then identified four-node motifs, the most significant of which are shown in Figure 1D. Among these, the forwarded uplinked mutual dyad forwards the output of an UMD to a further RBP and thus is a hierarchical, rank-connecting extension of this motif. Furthermore, the chain-feeding dyad is made of a dyad that transmits its regulatory signal to two linearly connected RBP mRNAs, thus creating 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 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. (2004). The TSP quantifies the use of the various three-node motifs by the network under analysis with respect to random networks 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 mRNA 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 predicted interactions between RBPs and RBPs mRNAs. This network, independently reconstructed from two experimental datasets, becomes a validation of the general structure we propose for the RBP-RBP network. It indeed includes all regions found to be bound by RBPs in HEK293 cells by an experimental approach, regions that were then assigned to a putative regulator by predictions based on experimentally determined RBP binding motifs.

We eventually compared the TSP of the three networks. The results are shown in Figure 1E, and we observe two salient aspects. First, the RBP-RBP network and its inferred version have a very similar motif structure (Spearman correlation = 0.78, p value = 2.62×10^{-3}), with limited magnitude differences only. This suggests that our network structure is reproducible and a representative cross-section of the complete set of interactions between RBPs and an RBP mRNA. Then, the TF-TF structure is instead more distant (mean Spearman correlation = 0.698 across the 41 networks). Indeed, 5/13 motifs are differentially

represented in the RBP-RBP network (enriched instead of depleted or vice versa), and the DMD is preferred over the UMD (the opposite being true for the TF-TF networks). This suggests specialization of network structures in RBP-RBP interactions with respect to TF-TF ones.

The Stoichiometry of RBP Complexes Is Not Determined by Interactions between RBPs and RBP mRNAs

We then asked if some type of biological constraints could be behind the evolutionary shaping of the specific geometry of the RBP-RBP network. One hypothesis is that constraints are produced by RBPs linked in the network being part of the same ribonucleoprotein complex, assuring coordinated post-transcriptional control to reach stoichiometry (i.e., are post-transcriptional interactions of an RBP and an RBP mRNA selected because of the need to regulate existent protein-protein interactions between those two RBPs?). To test this hypothesis, we overlapped the interactions in our experimental RBP-RBP network with the experimental binary protein-protein interactions (PPIs) contained in STRING (Szklarczyk et al., 2017), IntAct (Orchard et al., 2014), and BioPlex (Huttlin et al., 2017). We found only few RBP-RBP mRNA interactions mirrored by PPIs (3.37% for STRING, 0.57% for IntAct, and 0.29% for BioPlex). The same behavior, although to a slightly higher extent, appears for the TF-TF networks, with 3.27% for STRING, 1.30% for IntAct, and 0.43% for BioPlex in the Renal Cortical Epithelial Cell network. We then looked at the same data from the opposite perspective. In a related work based on yeast data, Pancaldi and Bähler (2011) observed that RBPs often bind to the mRNAs that encode their physical interaction partners. In our dataset, this behavior is infrequent, with only 3.8% and 10.7% of the RBP-RBP physical interactions from the STRING and IntAct databases (obtained by extracting all physical protein-protein interactions and filtering them by the involvement of two RBPs) respectively mirrored in our RBP-RBP network. Globally, this suggests that the network wiring is not generally made to assure the availability of RBPs for complex assembly.

As this analysis dealt with single interactions, we then turned to whole complexes. Indeed, stoichiometric complexes (i.e., requiring precise quantities of each of the components for proper functioning) may instead rely on this mechanism. We employed data from CORUM (Ruepp et al., 2010) and found 1818 interactions overlapping a complex, corresponding to only 3.79% of the network (for the TF-TF this figure is even lower, covering 2.53% of the Renal Cortical Epithelial Cell network). Table S3 lists complexes with at least five interactions in the network involving their subunits. A few complexes are highly represented, including the large Drosha complex (95% of its subunits are in the network, connected by 88 interactions) and the spliceosome (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. Furthermore, Pancaldi and Bähler, 2011 also hypothesized that RBPs in a complex may coordinate the expression of non-RBPs within the same complex. Of the 2,390 human protein complexes listed in CORUM, 658 contain at least one RBP and one non-RBP. In our dataset, only 35 of these (5.3%) have at least one non-RBP regulated by one RBP in the complex, with a mean of 58% of regulated non-RBPs and 13 (37%) having all non-RBPs regulated. Therefore, this observation does not hold in our network.

Communities Do Not Globally Define the Structure of the RBP-RBP Network

To obtain a more general understanding of post-transcriptional RBP-RBP interactions, we thus asked ourselves whether the network had a modular structure. Can we identify groups of densely connected RBPs (communities) potentially aimed at regulating specific biological processes? To answer this question, we used SurpriseMe (Aldecoa and Marin, 2013), a tool for the investigation of community structures. 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 = 13,698 and 13,353 resp.). These algorithms detected a poor degree of modularity in the network. As shown in Figure 2A, 89% of the communities are formed by a single RBP (singletons) and only eight contain more than 20 RBPs (one with CPM and seven 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 singletons (avg. of 53 vs. 657 for the RBP-RBP network) and higher 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

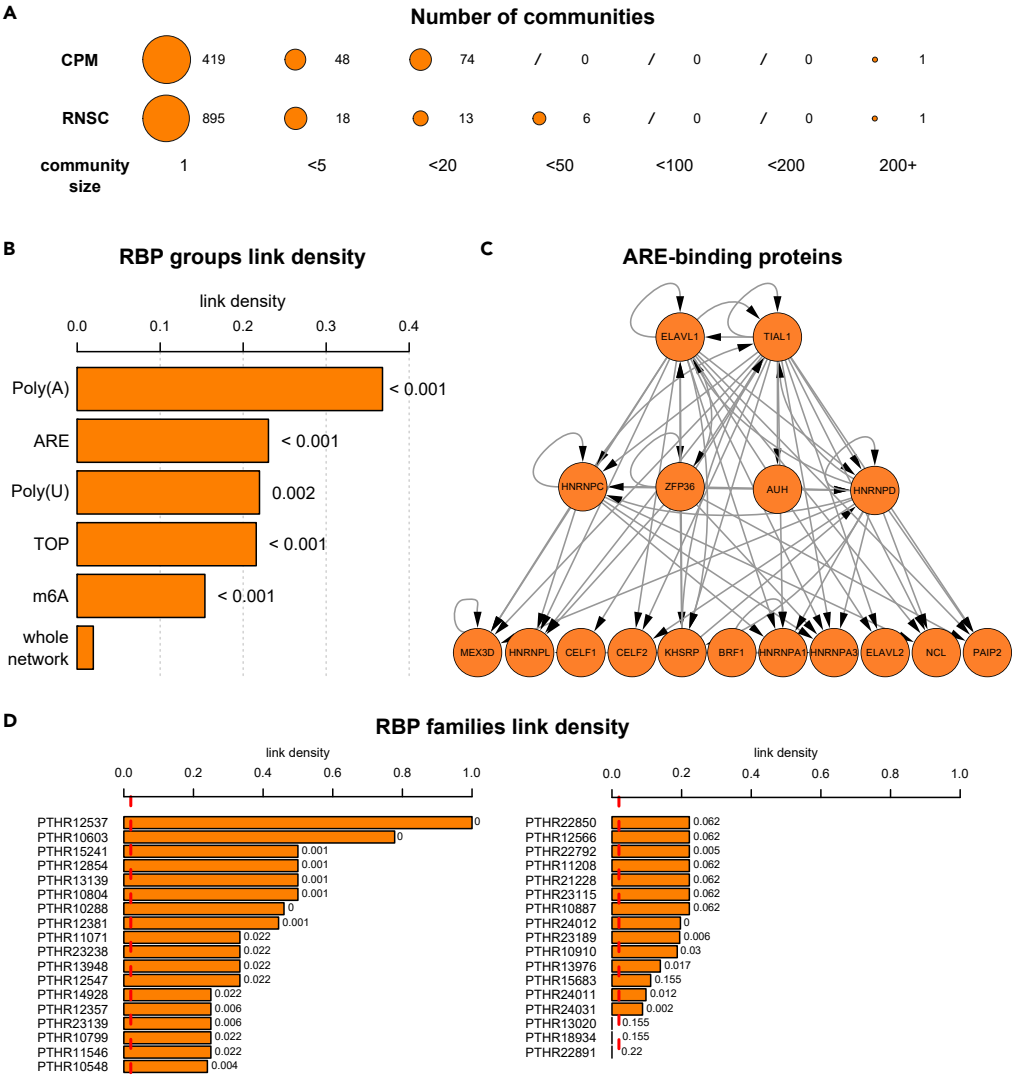


Figure 2. RBPs Binding Common Sets of Targets Are Significantly More Interconnected

(A) 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) 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) The complete network of ARE-binding proteins, revealing a hierarchical structure dominated by *ELAVL1* and *TIAL1*.

(D) 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.

See also Table S4.

listed in Tables 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.

RBP-RBP Interactions Occur in Clusters Dictated by Their Common Target mRNAs

The number and size of the detected communities suggest the low modularity of the RBP-RBP network, likely due to a peculiar community structure that cannot be detected by current algorithms. To further study this aspect we investigated a more general principle, that of interactions between RBPs in the network being due 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 mRNA targets for each RBP in the network

from the AURA 2 database (Dassi et al., 2014) and computed their 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 $p < 2.2 \times 10^{-16}$).

Examples of cooperative/competitive mechanisms that were already known, and for which the two participating RBPs bind to each other in our network, include *DDX3X* cooperating with *PABP1* (Copsey et al., 2017) and *CELF2* competing with *ELAVL1* (Sureban et al., 2007).

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 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), and proteins interacting with the m6A RNA modification sites (Roignant and Soller, 2017). We also include proteins interacting 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, with *HuR/ELAVL1* and *TIAL1* as top 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 may 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). An example of highly interacting family is the fragile X mental retardation, comprising *FMR1* and its two paralogs, *FXR1* and *FXR2*. These RBPs have largely overlapping targets and predominantly associate with polysomes to control brain development (Li and Zhao, 2014). In the network, *FMR1* binds to the other two RBPs, as does *FXR2*. All three also bind to their cognate mRNA. This suggests a potentially tight control over how these RBPs concur to regulate their common targets.

Globally, 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, these results further indicate that interactions between RBPs and RBP mRNAs may be needed to regulate cooperative and competitive behaviors on mutual targets. Furthermore, 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.

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. The latter were 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 of each pair of RBPs in a chain or community. Semantic similarity uses functional annotations (i.e., Gene Ontology) to define how close two genes are in terms of performed functions. 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 S1). Chains thus seem to be relevant to the RBP-RBP network organization. We performed this analysis also in the TF-TF networks, detecting chains of both length 4 and 5. These are more abundant than in the RBP-RBP network (1.5 times the longest chains in the Renal Cortical Ephitelial Cell network), but are significantly less functionally coherent (average

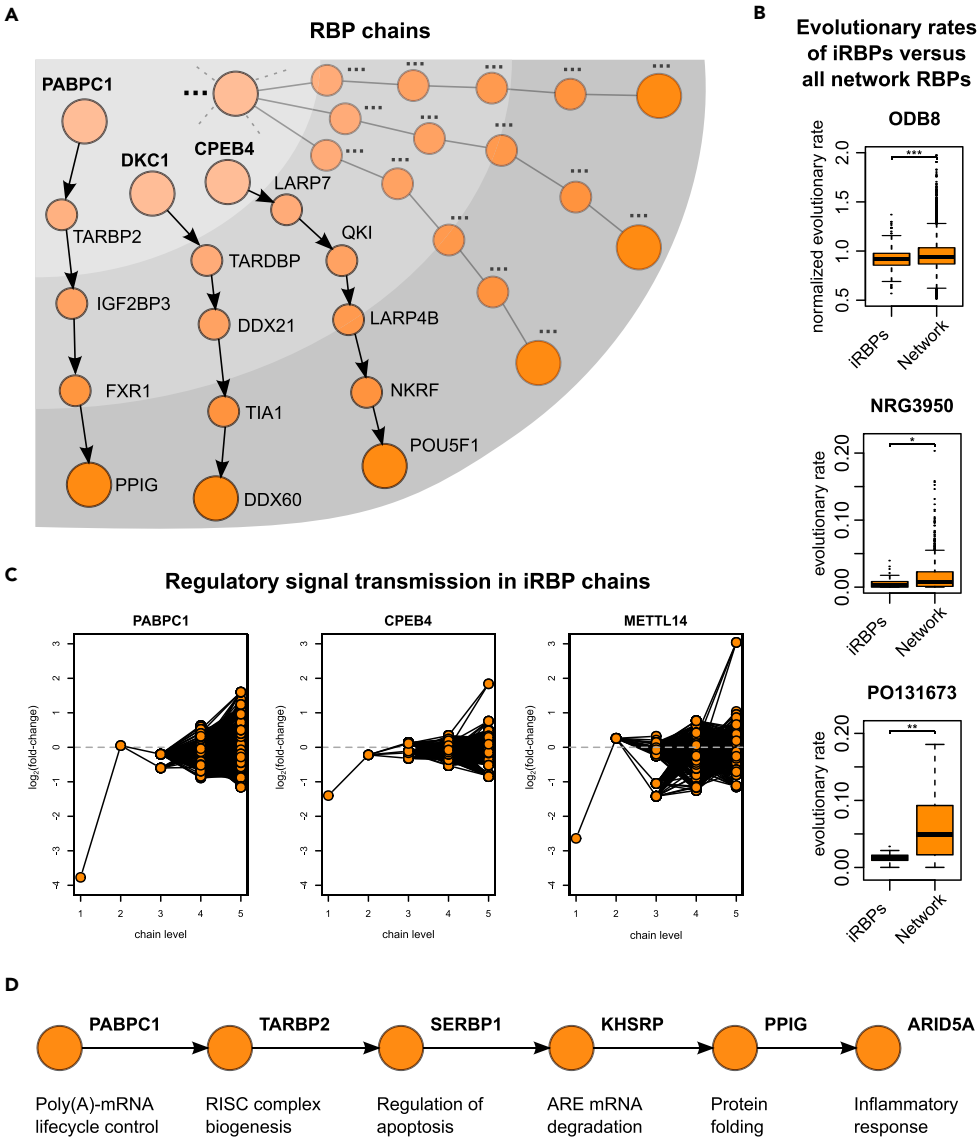


Figure 3. RBP Chains Dispatch Regulatory Information in the Network

(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 $p = 5.2 \times 10^{-5}$, 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, whereas level 5 represents the last RBP of a chain (with levels 2 and 4 being intermediate steps of each chain). Lines represent the RBP-RBP connections in a chain, whereas orange circles represent RBPs.

(D) An example of a chain led by PABPC1 and composed of RBPs performing a broad set of functions (main function shown only, as per Uniprot gene descriptions).

See also Figures S1–S3 and Tables S5–S8.

0.41/0.43 vs. 0.75/0.73 for TF-TF and RBP-RBP of length 4 and 5 resp.; Wilcoxon test $p < 2.2 \times 10^{-16}$ for both lengths), suggesting a lesser importance of such units in a denser network such as this one.

Chains are headed by a few initiator RBPs (iRBPs, 53 genes). iRBPs include genes involved in the regulation of various aspects of RNA metabolism and ribosome biogenesis. 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, one-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 $p = 1.73 \times 10^{-5}$), 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, *Drosophila melanogaster,* and *Caenorhabditis elegans*, and compared them with essential genes in those organisms. As shown in Tables S5–S7, the enrichment of essential genes in iRBPs is highly significant also for these organisms. Finally, if iRBPs are essential to cell physiology they should be rarely targeted by copy number imbalances in cancer, as their perturbation could negatively affect the cell fitness. Indeed, when analyzing 11,325 samples from 31 tumor types in TCGA, 7,656 samples (67.6%) have no CNA in any iRBP and 83.7% in no more than one iRBP. Of these CNAs, 75.5% are gains/amplifications (Figure S2).

We reasoned that the iRBPs could also be highly conserved, due to their fundamental role in driving 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 (Figure 3B; Wilcoxon test $p = 5.2 \times 10^{-5}$, 0.0128 and 0.0016 for ODB8, NRG3950, and PO131673, respectively). Furthermore, iRBP transcripts are significantly more expressed than other RBPs in HEK293 cells (median of 52.1 and 21.58 RPKM respectively, p value = 0.0065), in accordance with the observed association between evolutionary constraints and high expression (Gout et al., 2010). The essentiality of most iRBPs, coupled with their ultra-conservation, consistently support their importance as key cell regulators.

RBP Chains Transmit Post-transcriptional Signals

We eventually asked ourselves whether the potential regulatory information induced by RBP binding is transmitted through the chains, from the iRBPs down to the last node. To study this aspect, we reanalyzed publicly available transcriptome profiles of knock-down experiments for three iRBPs in human cells (two with chains of length 4 and 5, *PABPC1* and *CPEB4*, and one with chains of length 4, *METTL14,* see Methods). 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 sizable 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). When 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*; at this threshold, 49% of *PABPC1* chains have at least 80% of their composing RBPs as differentially expressed, with 52% for *CPEB4*, and 57% for *METTL14*). We also analyzed alternative splicing changes in these datasets and identified further activated chain links (i.e., that were not activated by differential gene expression but by differential exon usage). These involve 73 additional genes for *PABPC1* (15.3%), 54 for *CPEB4* (11.7%), and 40 for *METTL14* (15.6%). Globally, the total thus rises to 39.2% for *PABPC1*, 34.1% for *CPEB4*, and 61.9% for *METTL14* (adjusted p value threshold = 0.05). It must be noted that other modes of regulation, which cannot be observed through these datasets, can also be used by these proteins aside from mRNA stability and splicing (e.g., translational control). The amount of "activated" RBPs in the chains is thus here underestimated. This data suggest that the potentially regulatory information sparked by an iRBP is indeed transmitted through its chains. As a result, the set of processes that can be controlled by these proteins is likely broadened (an example of which is shown in Figure 3D). Consequently, we asked ourselves which pathways were affected by RBP chains. We thus overlapped the RBPs contained in each chain of length 5 with annotated human pathways from the MSigDB collection (Subramanian et al., 2005). Results, displayed in Figure S3, indicate that while globally the overlap with pathways is modest (only 124/1329—9.3% of annotated pathways include one or more chain RBP), several functions are represented. These include expected functions (e.g., RNA processing and metabolism, translation—overlapped by several chains) but also pathways related to less conventional functions of RBPs (such as DNA repair, immunity and viral

life cycle, intracellular signaling, and others, involving less chains). 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 form a hierarchical structure, as also suggested by the ranked clusters model we observed for 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 and of a random, degree-preserving network. 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 and a non-significant 0.99 for the random network at six levels. Furthermore, feedback loops (not coherent with a hierarchical organization) are depleted in the network, representing 0.0085% of the motifs only. Lastly, feedforward loops, coherent with a hierarchical organization, are instead enriched and amount to 3.29% of the motifs.

We then assessed another desirable property, that of network robustness to the "deletion" (i.e., loss of function) of an RBP from the network. To do so, we computed the pairwise disconnectivity metric on each RBP (Potapov et al., 2008). We found only 0.14% of RBP pairs to be disconnected on average when removing an RBP from the network (0.13% on a corresponding random, degree-preserving network), a significantly lower value than for the TF-TF networks (average is three times higher for TFs, worst p value = 5.6E-104). The network is thus well-tolerant to losing a node (on average, only a few RBPs are disconnected when "deleting" a given RBP), implying that RBP-RBP interactions are robust and suggesting the degree structure of the network as a contributing factor. This feature is likely granted by the use of densely connected RBP clusters, resulting in partially redundant binding to the same mRNAs.

Eventually, although RBP clusters are redundant by definition (as they co-regulate a largely overlapping set of targets), we asked whether single RBP chains also shared this property. We thus computed the overlap between all RBP targets (both RBPs in the network and non-RBPs outside it) at the various levels of each chain of length 5. It resulted being particularly low, as only 7.6% of the targets are overlapping between any two levels (median of all chains, average of each pair in a chain; the range is 2.8%–15.5%). Differently from RBP clusters, we can thus say that chains are efficient, as targets are not redundantly bound by individual RBPs along the chain. These are instead predominantly organized in complementary sets at each chain node. This efficiency comes at the expense of robustness (i.e., if one level of the chain fails the regulatory signal would most often be lost), which is instead a feature of RBP clusters. The resulting model, shown in Figure 4, couples hierarchical structure, network robustness through RBP clusters, and efficiency through RBP chains.

DISCUSSION

We presented here a systematic characterization of the RBP-RBP 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 the 5′ or 3′UTR of an RBP mRNA. The network is small-world and scale-free, typical properties of gene regulatory networks. Although the network is partial (as data are available for a fraction of all RBPs only), its local motif structure is highly coherent with the one of the inferred network. The network is thus representative of general patterns in RBP-RBP interactions. It must be noted that the network stems from the integration of various types of experimental techniques, not uniformly processed due to their different specific features. Furthermore, one should consider that not all interactions scored by the CLIP and RIP assays, from which the most part of the network is derived (i.e., an RBP binding to the UTRs of the mRNA of another RBP), are likely to produce an effect on the downstream levels and translation of the target transcript.

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. This specialization could be aimed at better suiting the specificities of post-transcriptional regulation, such as the need to rapidly react to cellular stress and stimuli. In particular, we found up- and down-linked mutual dyads as particularly enriched. These motifs are

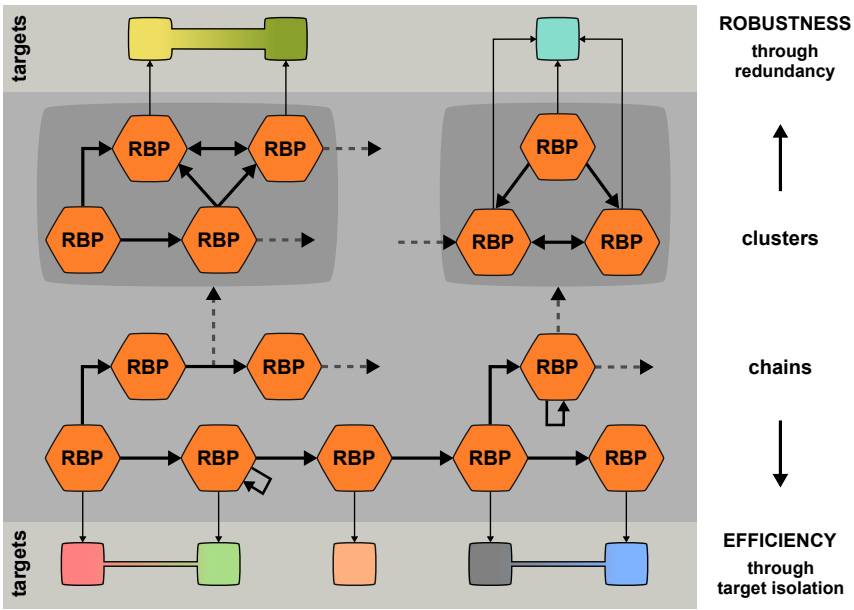


Figure 4. The RBP-RBP Network Is a Robust and Efficient Hierarchy

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 clusters (co-binding most of their targets), whereas 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 clusters and chains in the network.

distinctive of a ranked clusters structure (Johnsen, 1985), thus suggesting that the network can be divided into features conferring hierarchy and clusters of densely interacting nodes.

To study the role of these interactions in shaping cell phenotypes, we investigated why RBPs bind to each other mRNA. 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 clusters of RBPs having overlapping targets tend to bind to each other's mRNA. These interactions could represent a yet unappreciated 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 post-transcriptionally regulating the expression of the partner RBP. This mechanism could be used to reach precise ratios of mRNP components and yield the intended regulatory effect on common targets.

These clusters represent "islands" of densely connected RBPs and are key in providing robustness to the network. Indeed, their partially redundant binding could improve the resilience of the network to the loss-of-function of individual RBPs.

Surprisingly, the TF-TF network by Neph et al., 2012 appears much less tolerant to losing individual transcription factors than the RBP-RBP one (i.e., on average, much more nodes end up disconnected). We speculate that this could imply regulatory modules within the TF-TF network to be smaller and more isolated than for RBPs. The TF-TF structure could thus sustain the "fragmentation" of the network better. Another hypothesis may see the cost of redundantly wiring the TF-TF network as much higher than for the RBP-RBP one, thus making the toll of losing connections more bearable than that of having to replicate them. Further work may be warranted to investigate this aspect. Furthermore, these structural differences between TF-TF and RBP-RBP networks suggest the opportunity to integrate the two, potentially including also microRNA data, to study the combined effect and properties of the whole regulome. However, RBP clusters 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 the importance of iRBPs as ancient master regulators of cellular 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., targets are not replicated along a chain), thus possibly 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 clusters). 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-node 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 the fundamental cellular machinery driving the post-transcriptional regulation of gene expression. Establishing robustness only through RBP clusters could lead to a weaker architecture, as the regulatory signal going through chains is not redundant and a single perturbation of the network could cause its loss. However, it may be cheaper to obtain and more far-reaching. This consideration, however, raises a question: which is the role of RBP chains in relation to RBP clusters?

We suggest that RBP chains use the modulation of RBP targets as a connector to different processes, represented by the RBP clusters. Clusters would thus be similar in concept to the regulatory modules identified by Mukherjee et al., 2019, although made exclusively of RBPs regulating each other. We call this model the "*modular controller*." Under this model, the signal originated from the chains iRBPs proceeds from one RBP cluster, side-connected to the chain, to another, while flowing through the chain levels to control 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.

Limitations of the Study

It must be noted that the network is partial, as experimental data are currently available for a fraction of all RBPs only, and that we considered only UTR-mediated interactions (which however represent the biggest fraction of all known RBP-mRNA interactions). Furthermore, the network stems from the integration of various types of experimental techniques, not uniformly processed due to their different specific features. Eventually, one should also consider that not all interactions scored by the CLIP and RIP assays, from which the most part of the network is derived (i.e., an RBP binding to the UTRs of the mRNA of another RBP), are likely to be functional and thus produce an effect on the downstream levels and translation of the target transcript.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

An interactive browser allowing to explore the RBP-RBP network is available at the AURA 2 website (http://aura.science.unitn.it). The networks were deposited in NDEX (http://www.ndexbio.org) with ID *ee3e8898-6e29-11e8-a4bf-0ac135e8bacf*, *f5ad750b-6e29-11e8-a4bf-0ac135e8bacf*, and *fc1e526e-6e29-11e8-a4bf-0ac135e8bacf*.

The source scripts employed for non-trivial analyses described in the paper are available at https://bitbucket.org/erikdassi/rbp-rbp-network.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.10.058.

ACKNOWLEDGMENTS

This work was supported by core funding of the University of Trento (Trento, Italy).

AUTHOR CONTRIBUTIONS

E.D. designed the research. E.D. performed the analyses and prepared the figures. E.D. and A.Q. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: March 25, 2019
Revised: October 1, 2019
Accepted: October 28, 2019
Published: November 22, 2019

REFERENCES

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

Baltz, A.G., 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., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150.

Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092–1096.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Cheng, C., Andrews, E., Yan, K.-K., Ung, M., Wang, D., and Gerstein, M. (2015). An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. Genome Biol. *16*, 63.

Copsey, A.C., Cooper, S., Parker, R., Lineham, E., Lapworth, C., Jallad, D., Sweet, S., and Morley, S.J. (2017). The helicase, DDX3X, interacts with poly(A)-binding protein 1 (PABP1) and caprin-1 at the leading edge of migrating fibroblasts and is required for efficient cell spreading. Biochem. J. *474*, 3109–3120.

Dassi, E. (2017). Handshakes and fights: the regulatory interplay of RNA-binding proteins. Front. Mol. Biosci. *4*, 67.

Dassi, E., Zuccotti, P., Leo, S., Provenzani, A., Assfalg, M., D'Onofrio, M., Riva, P., and Quattrone, A. (2013). Hyper conserved elements in vertebrate mRNA 3′-UTRs reveal a translational network of RNA-binding proteins controlled by HuR. Nucleic Acids Res. *41*, 3201–3216.

Dassi, E., Re, A., Leo, S., Tebaldi, T., Pasini, L., Peroni, D., and Quattrone, A. (2014). AURA 2: empowering discovery of post-transcriptional networks. Translation (Austin) *2*, e27738.

de Nooy, W., Mrvar, A., and Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek (Cambridge University Press).

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Goss, D.J., and Kleiman, F.E. (2013). Poly(A) binding proteins: are they all created equal? Wiley Interdiscip. Rev. RNA *4*, 167–179.

Gout, J.-F., Kahn, D., and Duret, L.; Paramecium Post-Genomics Consortium (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet. *6*, e1000944.

Hamilton, T.L., Stoneley, M., Spriggs, K.A., and Bushell, M. (2006). TOPs and their regulation. Biochem. Soc. Trans. *34* (Pt 1), 12–16.

Humphries, M.D., and Gurney, K. (2008). Network "small-world-ness": a quantitative method for determining canonical network equivalence. PLoS One *3*, e0002051.

Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.

Hu, W., Yuan, B., and Lodish, H.F. (2014). Cpeb4-mediated translational regulatory circuitry controls terminal erythroid differentiation. Dev. Cell *30*, 660–672.

Johnsen, E.C. (1985). Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices. Soc. Network. *7*, 203–224.

Jothi, R., Balaji, S., Wuster, A., Grochow, J.A., Gsponer, J., Przytycka, T.M., Aravind, L., and Babu, M.M. (2009). Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. Mol. Syst. Biol. *5*, 294.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

King, A.D., Przulj, N., and Jurisica, I. (2004). Protein complex prediction via cost-based clustering. Bioinformatics *20*, 3013–3020.

Kleinberg, J.M. (2000). Navigation in a small world. Nature *406*, 845.

Kosti, I., Radivojac, P., and Mandel-Gutfreund, Y. (2012). An integrated regulatory network reveals pervasive cross-regulation among transcription and splicing factors. PLoS Comp. Biol. *8*, e1002603.

Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2015). Tissue-specific evolution of protein coding genes in human and mouse. PLoS One *10*, e0131673.

Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP technologies for studies of protein-RNA interactions. Mol. Cell *69*, 354–369.

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

Li, Y., and Zhao, X. (2014). Concise review: fragile X proteins in stem cell maintenance and differentiation. Stem Cells *32*, 1724–1733.

Liu, Q., Wang, Z., Li, M., Li, D., and 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. Available at: https://doi.org/10.1109/iccsit.2009.5234476.

Lukong, K.E., Chang, K.-W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet. *24*, 416–425.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nature reviews. Mol. Cell Biol. *8*, 479–490.

Mansfield, K.D., and Keene, J.D. (2009). The ribonome: a dominant force in co-ordinating gene expression. Biol. Cell *101*, 169–181.

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

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

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U., and 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.

Mukherjee, N., Wessels, H.-H., Lebedeva, S., Sajek, M., Ghanbari, M., Garzia, A., Munteanu, A., Yusuf, D., Farazi, T., Hoell, J.I., et al. (2019). Deciphering human ribonucleoprotein regulatory networks. Nucleic Acids Res. *47*, 570–581.

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

Newman, M.E.J., and Girvan, M. (2004). Finding and evaluating community structure in networks. Phys. Rev. E *69*, https://doi.org/10.1103/physreve.69.026113.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., 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* (Database issue), D358–D363.

Pancaldi, V., and Bähler, J. (2011). In silico characterization and prediction of global protein–mRNA interactions in yeast. Nucleic Acids Res. *39*, 5826–5836.

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

Pham, H., Ferrari, R., Cokus, S.J., Kurdistani, S.K., and Pellegrini, M. (2007). Modeling the regulatory network of histone acetylation in Saccharomyces cerevisiae. Mol. Syst. Biol. *3*, 153.

Potapov, A.P., Goemann, B., and 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, H.H., Abdelmohsen, K., Lal, A., Martindale, J.L., Yang, X., and Gorospe, M. (2007). Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. Mol. Cell. Biol. *27*, 6265–6278.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., 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., and 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., and Mewes, H.-W. (2010). CORUM: the comprehensive resource of mammalian protein complexes–2009. Nucleic Acids Res. *38* (Database issue), D497–D501.

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

Schmidt, E.E., Pelz, O., Buhlmann, S., Kerr, G., Horn, T., and Boutros, M. (2013). GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. Nucleic Acids Res. *41* (Database issue), D1021–D1026.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Sureban, S.M., Murmu, N., Rodriguez, P., May, R., Maheshwari, R., Dieckgraefe, B.K., Houchen, C.W., and Anant, S. (2007). Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. Gastroenterology *132*, 1055–1065.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45* (D1), D362–D368.

Tcherkezian, J., Cargnello, M., Romeo, Y., Huttlin, E.L., Lavoie, G., Gygi, S.P., and Roux, P.P. (2014). Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation. Genes Dev. *28*, 357–371.

Vogel, C., de Sousa Abreu, R., Ko, D., Le, S., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. *6*, 400.

Wang, Z., and Kiledjian, M. (2000). The poly(A)-binding protein and an mRNA stability protein jointly regulate an endoribonuclease activity. Mol. Cell. Biol. *20*, 6334–6341.

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., et al. (2018). METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA mA modification. Cell Stem Cell *22*, 191–205.e9.

Wernicke, S., and Rasche, F. (2006). FANMOD: a tool for fast network motif detection. Bioinformatics *22*, 1152–1153.

Wurth, L., and Gebauer, F. (2015). RNA-binding proteins, multifaceted translational regulators in cancer. Biochim. Biophys. Acta *1849*, 881–886.

Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. *45* (D1), D744–D749.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.

Zhang, J., and Yang, J.-R. (2015). Determinants of the rate of protein sequence evolution. Nat. Rev. Genet. *16*, 409–420.

ISCI, Volume *21*

Supplemental Information

The Architecture of the Human RNA-Binding Protein Regulatory Network

Alessandro Quattrone and Erik Dassi

Supplemental Figures

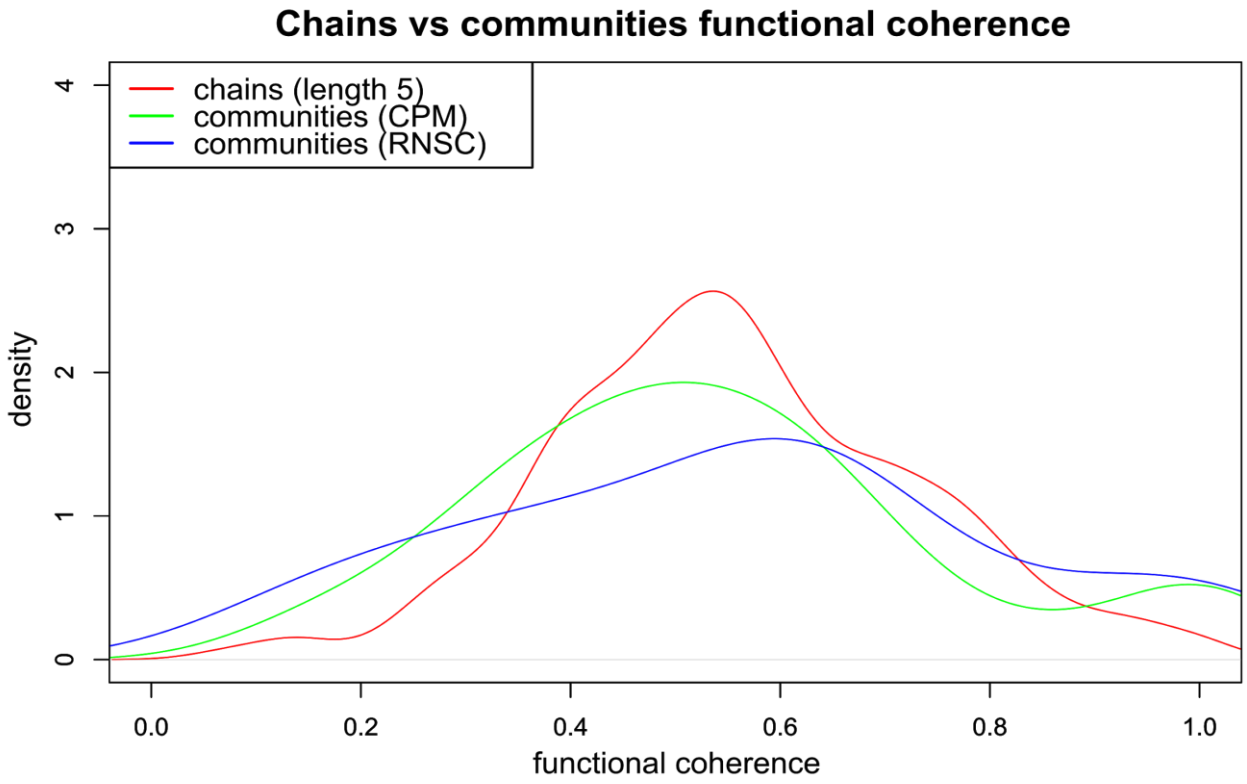


Figure S1: Functional coherence distribution for RBP chains versus communities. Related to Figure 3. Shows the differences in density of the functional coherence values distribution between RBP chains of length 5 (in red) and communities obtained by CPM (green) and RNSC (blue).

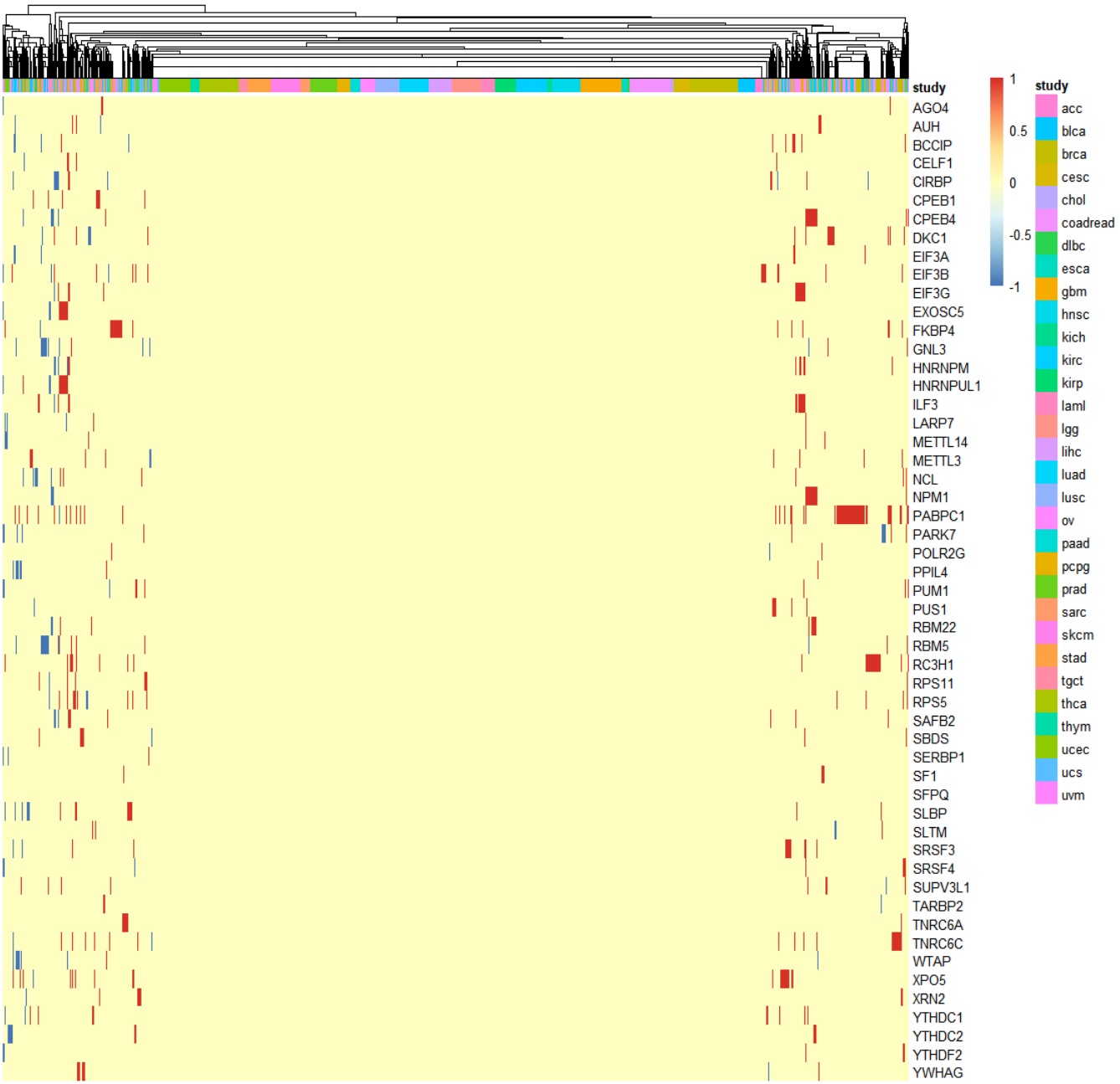


Figure S2: Copy number imbalances of iRBPs in TCGA cancer samples. Related to Figure 3.

Displays copy number alterations of iRBPs in 11325 samples from 31 tumor types of the TCGA dataset. Deletions are shown in blue while gains/amplifications are marked in red.

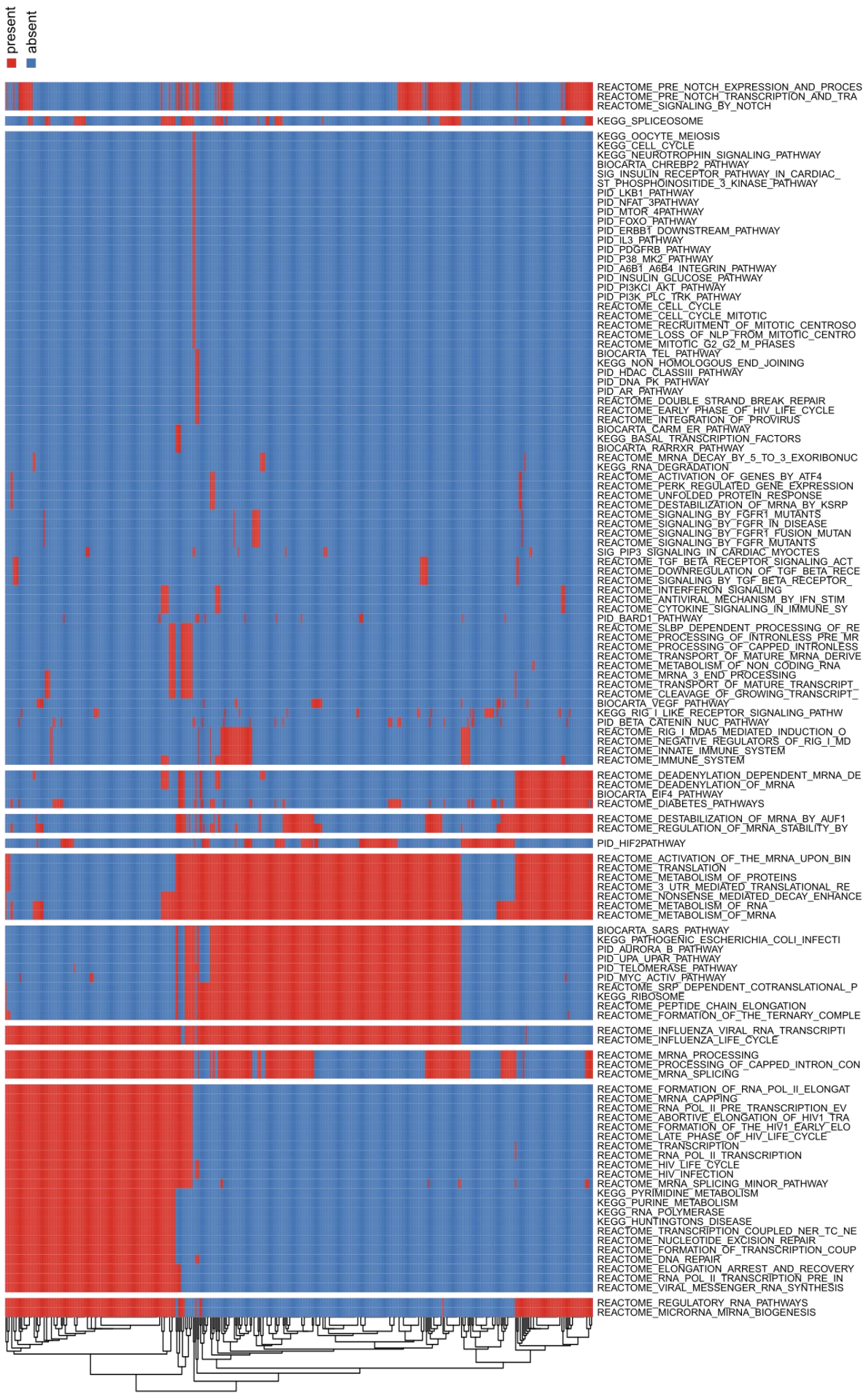

present
absent
REACTOME_PRE_NOTCH_EXPRESSION_AND_PROCES
REACTOME_PRE_NOTCH_TRANSCRIPTION_AND_TRA
REACTOME_SIGNALING_BY_NOTCH
KEGG_SPLICEOSOME
KEGG_OOCYTE_MEIOSIS
KEGG_CELL_CYCLE
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY
BIOCARTA_CHREBP2_PATHWAY
SIG_INSULIN_RECEPTOR_PATHWAY_IN_CARDIAC_
ST_PHOSPHOINOSITIDE_3_KINASE_PATHWAY
PID_LKB1_PATHWAY
PID_NFAT_3PATHWAY
PID_MTOR_4PATHWAY
PID_FOXO_PATHWAY
PID_ERBB1_DOWNSTREAM_PATHWAY
PID_IL3_PATHWAY
PID_PDGFRB_PATHWAY
PID_P38_MK2_PATHWAY
PID_A6B1_A6B4_INTEGRIN_PATHWAY
PID_INSULIN_GLUCOSE_PATHWAY
PID_PI3KCI_AKT_PATHWAY
PID_PI3K_PLC_TRK_PATHWAY
REACTOME_CELL_CYCLE
REACTOME_CELL_CYCLE_MITOTIC
REACTOME_RECRUITMENT_OF_MITOTIC_CENTROSO
REACTOME_LOSS_OF_NLP_FROM_MITOTIC_CENTRO
REACTOME_MITOTIC_G2_G2_M_PHASES
BIOCARTA_TEL_PATHWAY
KEGG_NON_HOMOLOGOUS_END_JOINING
PID_HDAC_CLASSIII_PATHWAY
PID_DNA_PK_PATHWAY
PID_AR_PATHWAY
REACTOME_DOUBLE_STRAND_BREAK_REPAIR
REACTOME_EARLY_PHASE_OF_HIV_LIFE_CYCLE
REACTOME_INTEGRATION_OF_PROVIRUS
BIOCARTA_CARM_ER_PATHWAY
KEGG_BASAL_TRANSCRIPTION_FACTORS
BIOCARTA_RARRXR_PATHWAY
REACTOME_MRNA_DECAY_BY_5_TO_3_EXORIBONUC
KEGG_RNA_DEGRADATION
REACTOME_ACTIVATION_OF_GENES_BY_ATF4
REACTOME_PERK_REGULATED_GENE_EXPRESSION
REACTOME_UNFOLDED_PROTEIN_RESPONSE
REACTOME_DESTABILIZATION_OF_MRNA_BY_KSRP
REACTOME_SIGNALING_BY_FGFR1_MUTANTS
REACTOME_SIGNALING_BY_FGFR_IN_DISEASE
REACTOME_SIGNALING_BY_FGFR1_FUSION_MUTAN
REACTOME_SIGNALING_BY_FGFR_MUTANTS
SIG_PIP3_SIGNALING_IN_CARDIAC_MYOCTES
REACTOME_TGF_BETA_RECEPTOR_SIGNALING_ACT
REACTOME_DOWNREGULATION_OF_TGF_BETA_RECE
REACTOME_SIGNALING_BY_TGF_BETA_RECEPTOR_
REACTOME_INTERFERON_SIGNALING
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIM
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SY
PID_BARD1_PATHWAY
REACTOME_SLBP_DEPENDENT_PROCESSING_OF_RE
REACTOME_PROCESSING_OF_INTRONLESS_PRE_MR
REACTOME_PROCESSING_OF_CAPPED_INTRONLESS
REACTOME_TRANSPORT_OF_MATURE_MRNA_DERIVE
REACTOME_METABOLISM_OF_NON_CODING_RNA
REACTOME_MRNA_3_END_PROCESSING
REACTOME_TRANSPORT_OF_MATURE_TRANSCRIPT_
REACTOME_CLEAVAGE_OF_GROWING_TRANSCRIPT_
BIOCARTA_VEGF_PATHWAY
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHW
PID_BETA_CATENIN_NUC_PATHWAY
REACTOME_RIG_I_MDA5_MEDIATED_INDUCTION_O
REACTOME_NEGATIVE_REGULATORS_OF_RIG_I_MD
REACTOME_INNATE_IMMUNE_SYSTEM
REACTOME_IMMUNE_SYSTEM
REACTOME_DEADENYLATION_DEPENDENT_MRNA_DE
REACTOME_DEADENYLATION_OF_MRNA
BIOCARTA_EIF4_PATHWAY
REACTOME_DIABETES_PATHWAYS
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1
REACTOME_REGULATION_OF_MRNA_STABILITY_BY
PID_HIF2PATHWAY
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BIN
REACTOME_TRANSLATION
REACTOME_METABOLISM_OF_PROTEINS
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_RE
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCE
REACTOME_METABOLISM_OF_RNA
REACTOME_METABOLISM_OF_MRNA
BIOCARTA_SARS_PATHWAY
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTI
PID_AURORA_B_PATHWAY
PID_UPA_UPAR_PATHWAY
PID_TELOMERASE_PATHWAY
PID_MYC_ACTIV_PATHWAY
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_P
KEGG_RIBOSOME
REACTOME_PEPTIDE_CHAIN_ELONGATION
REACTOME_FORMATION_OF_THE_TERNARY_COMPLE
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTI
REACTOME_INFLUENZA_LIFE_CYCLE
REACTOME_MRNA_PROCESSING
REACTOME_PROCESSING_OF_CAPPED_INTRON_CON
REACTOME_MRNA_SPLICING
REACTOME_FORMATION_OF_RNA_POL_II_ELONGAT
REACTOME_MRNA_CAPPING
REACTOME_RNA_POL_II_PRE_TRANSCRIPTION_EV
REACTOME_ABORTIVE_ELONGATION_OF_HIV1_TRA
REACTOME_FORMATION_OF_THE_HIV1_EARLY_ELO
REACTOME_LATE_PHASE_OF_HIV_LIFE_CYCLE
REACTOME_TRANSCRIPTION
REACTOME_RNA_POL_II_TRANSCRIPTION
REACTOME_HIV_LIFE_CYCLE
REACTOME_HIV_INFECTION
REACTOME_MRNA_SPLICING_MINOR_PATHWAY
KEGG_PYRIMIDINE_METABOLISM
KEGG_PURINE_METABOLISM
KEGG_RNA_POLYMERASE
KEGG_HUNTINGTONS_DISEASE
REACTOME_TRANSCRIPTION_COUPLED_NER_TC_NE
REACTOME_NUCLEOTIDE_EXCISION_REPAIR
REACTOME_FORMATION_OF_TRANSCRIPTION_COUP
REACTOME_DNA_REPAIR
REACTOME_ELONGATION_ARREST_AND_RECOVERY
REACTOME_RNA_POL_II_TRANSCRIPTION_PRE_IN
REACTOME_VIRAL_MESSENGER_RNA_SYNTHESIS
REACTOME_REGULATORY_RNA_PATHWAYS
REACTOME_MICRORNA_MIRNA_BIOGENESIS

Figure S3: Association of chain RBPs to human pathways. Related to Figure 3. Shows the presence of RBPs belonging to a given chain (rows) in annotated human pathways from different sources (columns; MSigDB collection – see Methods). A red square indicated that at least one RBP of that chain is annotated to that specific pathway, a blue square that none are.

Supplemental Tables

Property	HEK293	HeLa	MCF7
Network diameter	5	5	6
Clustering coefficient (CC1)	0.507	0.5061	0.4938
Avg. clustering coefficient (CC2)	0.0115	0.0115	0.0108
Avg. closeness centrality	0.5033	0.5035	0.5013
S^{ws}	31.03	30.64	31
Control profile	s=0.00368, e=0.99632, i=0.0	s=0.00450, e=0.99474, i=0.00075	s=0.00370, e=0.99630, i=0.0

Table S2: Global properties of the RBP regulatory network in the three cell-lines. Related to Figure 1. This table lists the values of some commonly used global network properties for the RBP regulatory network in HEK293, HeLa and MCF7 cell-lines. These values show that the network structure is stable along this three systems.

RBP	Chain length	# of reachable RBPs	# of chains	Essential (Blomen_KBM7)	Essential (Blomen_HAP1)	Essential (GenomeRNAi)
AGO4	4	64	304			
AUH	4	27	45			v
BCCIP	4	89	542	v	v	v
CELF1	4	234	3808		v	v
CIRBP	4	220	4531			v
CPEB1	4	124	2461		v	v
CPEB4	4	461	10951			v
	5	26	47			
DKC1	4	112	1032	v	v	v

EIF3A	4	29	66	v	v	v
EIF3B	4	27	47	v	v	v
EIF3G	4	29	68		v	v
EXOSC5	4	37	102	v	v	v
FKBP4	4	60	155			v
GNL3	4	66	249	v	v	v
HNRNPM	4	52	167	v	v	v
HNRNPUL1	4	56	135			
ILF3	4	29	42	v	v	v
LARP7	4	25	47			
METTL14	4	257	5058	v	v	
METTL3	4	120	1531	v	v	v
NCL	4	516	14362	v	v	v
	5	63	202			
NPM1	4	57	437	v	v	v
PABPC1	4	481	4443			v
	5	55	67	v	v	
PARK7	4	233	3247			v
POLR2G	4	951	11195		v	v
	5	66	129			
PPIL4	4	83	312	v	v	v
PUM1	4	35	99	v	v	v
PUS1	4	40	76			v
RBM22	4	32	85	v	v	v
RBM5	4	48	89		v	v
RC3H1	4	61	231			v
RPS11	4	28	67	v	v	v
RPS5	4	62	202			v
SAFB2	4	72	314			v
SBDS	4	64	184	v	v	
SERBP1	4	28	47	v	v	v
SF1	4	170	2503	v	v	
SFPQ	4	65	129	v	v	v
SLBP	4	280	6937	v	v	v
SLTM	4	45	137	v		
SRSF3	4	228	3790	v	v	v
SRSF4	4	228	3790			
SUPV3L1	4	76	241	v		v
TARBP2	4	54	67			v
TNRC6A	4	74	273		v	
TNRC6C	4	74	189			
WTAP	4	200	4889			v
XPO5	4	33	108	v	v	v
XRN2	4	39	109	v	v	v
YTHDC1	4	138	2490	v		v
YTHDC2	4	88	309			v
YTHDF2	4	54	121			v

YWHAG	4	40	90			v

Table S5: iRBPs. Related to Figure 3. The table lists initiator RBPs (iRBPs) for regulatory chains of length 4 and 5. Listed are the number of chains headed by each RBP, the number of reached RBPs and the essentiality in two human cell lines as per (Blomen et al. 2015) and in RNAi screenings from GenomeRNAi (Schmidt et al. 2013). iRBPs are enriched in essential genes and a significant fraction (30%, bootstrap $p<0.001$) of these is essential in both cell lines. Merging the three sources yields 46/53 iRBPs (86%) as essential.

RBP	Essential
Ago4	
Auh	
Bccip	
Celf1	
Cirbp	
Cpeb1	
Cpeb4	
Dkc1	v
Eif3a	
Eif3b	v
Eif3g	
Exosc5	
Fkbp4	
Gnl3	v
Hnrnpm	
Hnrnpul1	
Ilf3	
Larp7	
Mettl14	
Mettl3	
Ncl	
Pabpc2	
Park7	
Polr2g	
Ppil4	
Pum1	
Pus1	
Rbm22	
Rbm5	
Rc3h1	

Rps11	
Rps5	
Safb2	
Sbds	v
Serbp1	
Sf1	v
Sfpq	
Slbp	
Sltm	
Srsf3	v
Srsf4	
Supv3l1	v
Tarbp2	
Tnrc6a	v
Tnrc6c	
Wtap	v
Xpo5	
Xrn2	
Ythdc1	
Ythdc2	
Ythdf2	
Ywhag	

Table S6: Essential master RBPs in Mus musculus. Related to Figure 3. Essential mouse orthologs of master RBPs (9/52) were derived as the ones conferring an embryonic lethality phenotype in mouse, as per the Phenotype Ontology of the MGI mouse database (www.informatics.jax.org).

1000-samples bootstrap p-value: 0.007

Enrichment p-value (Fisher test): 0.0039

RBP	**Essential**
ago1	
14-3-3zeta	v
B52	v
bgcn	
bru1	
bru2	
CG14641	
CG30122	

CG4887	
CG4896	
CG5808	v
CG8549	
CG8778	
CG9286	
DJ-1alpha	v
dj-1beta	
eIF3a	
eIF3b	
eIF3g1	
eIF3g2	
FDY	
FKBP59	v
fl(2)d	v
gw	
Ime4	
Larp7	
loqs	v
Mettl14	
nonA	v
nonA-l	v
Nop60B	v
Ns1	v
orb	v
orb2	v
pum	v
PUS1	
Ranbp21	
Rat1	v
roq	
Rpb7	v
RpS11	v
RpS5a	v
RpS5b	v
Rrp46	
rump	v
Saf-B	
SF1	v
Slbp	v
Suv3	
tra2	v
vig	
vig2	v
x16	v
YT521-B	

Table S7: Essential master RBPs in Drosophila melanogaster. Related to Figure 3. Essential *D.melanogaster* orthologs of master RBPs (25/54) were derived as the ones conferring a lethality phenotype in *D.melanogaster*, as per the Phenotype Ontology of the FlyBase database (flybase.org).

1000-samples bootstrap p-value: <0.001

Enrichment p-value (Fisher test): 0.032

RBP	Essential
alg-1	v
alg-2	v
cdl-1	v
cpb-3	v
crn-5	v
djr-1.1	
djr-1.2	
ech-5	
egl-45	v
eif-3.B	v
eif-3.G	v
etr-1	v
fbf-1	v
fbf-2	v
fkb-6	v
ftt-2	v
hrpu-1	
nono-1	v
nst-1	v
par-5	v
puf-11	v
puf-3	v
puf-5	v
puf-6	v
puf-7	v
pus-1	
rbm-22	
rbm-5	
rle-1	v
rpb-7	v
rps-11	v
rps-5	v
rsp-1	v

rsp-2	v
rsp-5	v
rsp-6	v
rsp-8	v
sbds-1	
sfa-1	v
sig-7	v
xrn-2	v

Table S8: Essential master RBPs in C.elegans. Related to Figure 3. Essential *C.elegans* orthologs of master RBPs (33/41) were derived as the ones conferring a lethality phenotype in *C.elegans*, as per the Phenotype Ontology of the WormBase database (wormbase.org).

1000-samples bootstrap p-value: <0.001

Enrichment p-value (Fisher test): <2.2E-16

Transparent 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 being protein-coding from Ensembl v92 (Zerbino et al. 2018), then merging these genes with the curated RBP list from (Sebestyén et al. 2016). The resulting catalogue 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 those made of more than one RBP.

Network construction

Regulatory interactions involving an RBP and the mRNA of an RBP were extracted from the AURA 2 database v2.4.4 (Dassi et al. 2014). The database is populated with significant RBP-mRNA UTR interactions (adjusted p-value <=0.05, determined by the specific analysis pipeline of each data type) derived from high-throughput (e.g. RIP-seq, CLIP-seq, and variants) as well as low-throughput assays (e.g. RIP-qPCR). Interactions were filtered by requiring the expression of both participants

(the regulating RBP and the target gene mRNA) in HEK293, HeLa or MCF7 cells, systems from which the majority of the data were derived. Expressed genes and related expression levels were determined by using RNA-seq profiles of HEK293 (Kishore et al. 2011), HeLa (Cabili et al. 2011) and MCF7 (Vanderkraats et al. 2013) cells, using an expression threshold of 0.1 RPKM. Each RBP/RBP mRNA is represented by a node, and edges represent regulation by the source RBP on the target RBP mRNA, with the edge direction defining the regulator (source node) and the regulated mRNA (target node).

To build the inferred RBP-RBP network, we employed RBP-bound regions in mRNA UTRs from a protein occupancy profiling assay (POP) performed in HEK293 cells (Baltz et al. 2012). Briefly, significant regions as defined in this work (i.e. mRNA parts detected to be bound by an RBP) were extracted and filtered for being within a 5' or 3'UTR according to Ensembl v92 annotation. Then, RNAcompete-derived PWMs, describing the binding motif of 193 human RBPs (Ray et al. 2013), were obtained from CISBP-RNA (Ray et al. 2013). Putative binding positions of these 193 RBPs in POP protein-bound regions were identified by the Biopython Bio.motifs module (Cock et al. 2009), using a relative score threshold of 0.99. We assigned the RBP with the highest score in each region as the one most likely to bind to that region, thus obtaining an (RBP, bound mRNA) pair for each UTR portion analyzed by Biopython. Eventually, only interactions involving two RBPs (the first binding to a POP region in the UTRs of the second) were used to build the network. Nodes and edges were defined as described for the experimental network

Network properties analysis

Network diameter, degree distribution, closeness centrality, Watts-Strogatz (CC1) and two-neighbor (CC2) clustering coefficient were computed by Pajek 3.14 (Batagelj & Mrvar 2002) and plotted with R (Tierney 2012). The S^{WS} measure, quantifying the likelihood of the network being actually small-world, was computed as described in (Humphries & Gurney 2008) by using the Watts-Strogatz clustering coefficient and generating the required random network with Pajek 3.14 (Batagelj & Mrvar 2002). The network control structure was computed by Zen 0.9 (Ruths & Ruths 2014) with default parameters. The hierarchical score was computed using the source code provided in the paper

defining this measure (Cheng et al. 2015), and pairwise disconnectivity obtained by the DiVa software (Potapov et al. 2008).

The link density of a set of nodes was computed as (number of links between nodes in the set) / (number of nodes in the set^2). Bootstraps were performed by 1000 random selections of a number of nodes equal to the set size and computation of the link density for each of these.

Network structure analysis

Network motifs of size 3 and 4 were identified with FANMOD (Wernicke & Rasche 2006) using 1000 random networks (100 for motifs of size 4, due to required computing time), 3 exchanges per edge and 3 exchange attempts. Triad significance profiles (describing the enrichment/depletion of each possible motif of a given size) for motifs of size 3 were computed as described in (Milo et al. 2004) for the RBP-RBP network, the inferred RBP-RBP network and the TF-TF networks described in (Neph et al. 2012).

Communities were detected with the SurpriseMe 1.0.4 tool (Aldecoa & Marin 2013), employing several community-detection tools and summarizing their result by the S surprise measure. CPM (Palla et al. 2005) and RNSC (King et al. 2004), the algorithms obtaining the highest S values, were eventually used to define communities. Chains of length 3, 4, and 5 were extracted from the network with igraph 0.7 (http://igraph.org); functional coherence scores were computed with GOSemSim 2.8 (Yu et al. 2010) as the average semantic similarity (defined over Gene Ontology functional annotations) of each pair of genes in a chain or community.

Protein-protein interactions, complexes, and pathways overlap

Human protein-protein interactions were extracted from STRING (Szklarczyk et al. 2017), BioPlex (Huttlin et al. 2017), and IntAct (Orchard et al. 2014), retaining only interactions of the “binding” type (physical association) and with both proteins involved being in our network. Human protein complexes were downloaded from the CORUM database (Ruepp et al. 2010). Overlaps were performed by using base library Python functions. RBP chains were overlapped with pathways derived from the MSigDB v7.0 collection (Subramanian et al. 2005).

Gene essentiality and phylogenetic conservation analysis

Essential genes of human cells were obtained from (Blomen et al. 2015); genes associated with an embryonic lethal phenotype in mouse from the MGI (Blake et al. 2017); genes associated with a lethality phenotype were extracted from WormBase (Lee et al. 2018) and FlyBase (Gramates et al. 2017) for C.*elegans* and D.*melanogaster* respectively. Orthologs of iRBPs were also extracted from the same databases. Bootstraps were computed by 1000 random selections of as many genes as iRBPs and computing the fraction of these in the essential genes set of each organism, thus determining an empirical p-value. The copy number alteration status for each iRBP in TCGA tumor samples was obtained from the cBIO portal v3.1 (Gao et al. 2013).

UTR conservation scores were computed by averaging the nucleotide-resolution phastCons scores derived from the UCSC 46-way vertebrate alignment (Casper et al. 2018). The average score of all 5' or 3' UTRs of a gene was employed as the conservation score for that gene 5' or 3'UTRs. Protein evolutionary rates were obtained from the ODB8 database (Zdobnov et al. 2017) and two published works (Kryuchkova-Mostacci & Robinson-Rechavi 2015; Zhang & Yang 2015); Wilcoxon tests were performed by R (Tierney 2012).

iRBP knock-down datasets

RNA-seq datasets following the knock-down of PABPC1, CPEB4, and METTL14 were obtained from GEO (IDs: GSE88099, GSE88545, and GSE56010). Reads were quality-trimmed (min quality score Q30, no Ns in the read, minimum read length after processing 36nts) and adapters removed with Trimmomatic (Bolger et al. 2014), then aligned to the human genome (GENCODE hg38 assembly), and transcripts quantified (Gencode v28 annotation) with STAR 2.5.2b (Dobin et al. 2013). Differential expression was eventually computed with DESeq2 v1.22.2 (Love et al. 2014) using default parameters and an adjusted p-value threshold of 0.05.

Supplemental References

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

Blake, J.A. et al., 2017. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic acids research*, 45(D1), pp.D723–D729.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* , 30(15), pp.2114–2120.

Cabili, M.N. et al., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*, 25(18), pp.1915–1927.

Casper, J. et al., 2018. The UCSC Genome Browser database: 2018 update. *Nucleic acids research*, 46(D1), pp.D762–D769.

Cock, P.J.A. et al., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* , 25(11), pp.1422–1423.

Dobin, A. et al., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , 29(1), pp.15–21.

Gao, J. et al., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, 6(269), p.l1.

Gramates, L.S. et al., 2017. FlyBase at 25: looking to the future. *Nucleic acids research*, 45(D1), pp.D663–D671.

Humphries, M.D. & Gurney, K., 2008. Network “small-world-ness”: a quantitative method for determining canonical network equivalence. *PloS one*, 3(4), p.e0002051.

Kishore, S. et al., 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature methods*, 8(7), pp.559–564.

Lee, R.Y.N. et al., 2018. WormBase 2017: molting into a new stage. *Nucleic acids research*,

46(D1), pp.D869–D874.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), p.550.

Sebestyén, E. et al., 2016. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome research*, 26(6), pp.732–744.

Tierney, L., 2012. The R Statistical Computing Environment. In *Lecture Notes in Statistics*. pp. 435–447.

Vanderkraats, N.D. et al., 2013. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic acids research*, 41(14), pp.6816–6827.

Yu, G. et al., 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* , 26(7), pp.976–978.

Zerbino, D.R. et al., 2018. Ensembl 2018. *Nucleic acids research*, 46(D1), pp.D754–D761.