The architecture of the human RNA-binding protein regulatory network

Erik Dassi, Paola Zuccotti, Daniele Peroni, Valentina Potrich and Alessandro Quattrone

Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Trento, Italy

Abstract

RNA-binding proteins (RBPs) are key players in post-transcriptional regulation of gene expression, implicated in both cellular physiology and pathology. Several studies have described individual interactions of RBP proteins with RBP mRNAs, evocative of an RBP regulatory hierarchy. Here we present the first systematic investigation of this hierarchy, based on a network including twenty thousand experimentally determined interactions between RBPs and bound RBP mRNAs. RBPs bind their mRNA in half of the cases, providing conclusive evidence of their general propensity to autoregulation. The RBP regulatory network is dominated by directional chains, rather than by modular communities as in the transcription factor regulatory network. These chains are initiated by an essential class of RBPs, the iRBPs. One of these, RBMX, initiates several thousand chains reaching most human RBPs. Our results show that chains make the network hierarchical, forming a post-transcriptional backbone that allows the fine-grained control of RBPs and their targets.

Introduction

Post-transcriptional regulation of gene expression has progressively gained recognition as a crucial determinant of protein levels and consequent cell phenotypes (Schwanhausser, Busse et al., Vogel, Abreu Rde et al.): this has resulted in a rising interest in studies focused on RNA-binding proteins (RBPs) and the interactions with their RNA targets.

RBPs are a key class of players in this regulatory layer, involved in controlling processes ranging from splicing and polyadenylation to mRNA localization, stability, and translation. RBPs

are less than two thousand proteins in the human genome (almost 1200 verified RBPs plus several recently discovered ones (Castello, Fischer et al.)). They are made of modular domains of which RRM is the most represented one, found in more than 200 RBPs (Lunde, Moore et al.). Techniques such as ribonucleoprotein immunoprecipitation (RIP) and cross-linking and immunoprecipitation (CLIP) variants (Milek, Wyler et al.) now allow us to massively identify the RNA targets of an RBP. RBPs are involved in multiple aspects of cell physiology (e.g. brain and ovary development, immune response and the circadian cycle (Gerstberger, Hafner et al., Lim & Allada)) and pathology, being their alteration associated with a variety of diseases such as cancer, neurological disorders, and muscular atrophies (Castello, Fischer et al., Lukong, Chang et al.). The importance of gaining a proper understanding of RBP properties and functions is thus evident.

While identifying the mRNA targets of RBPs, several works have highlighted an enrichment of mRNAs coding for gene expression regulators, including other RBPs but also transcription factors (TFs). This brought to the *regulator-of-regulators* concept (Keene, Mansfield & Keene), hinting at the existence of an extensive regulatory hierarchy of RBPs. For example, for this feature we and others have specifically studied the *HuR*/*ELAVL1* protein (Dassi, Zuccotti et al., Mukherjee, Corcoran et al., Pullmann, Kim et al.), which resulted to regulate the mRNAs of many RBPs (Mukherjee et al.), several of which contain its same RNA-binding domain, the RRM (Dassi et al.). The increasing number of high-throughput assays available is now allowing us to observe this phenomenon on a genome-wide scale.

We chose to address this issue by tracing a network of RBPs binding and regulating their cognate mRNA and other RBPs mRNA. This approach has been applied for TF targets and metabolic networks in lower organisms such as *E. coli* and *S. cerevisiae* (Hu, Killion et al., Jothi, Balaji et al., Pham, Ferrari et al.); the human TF-TF regulatory interaction network, applying the *regulator-of-regulators* concept to TFs, has also been described in 41 cell types (Neph, Stergachis et al.).

We present here the first systematic characterization of the RBP regulatory network, built by integrating experimental data on RBP targets. The network is scale-free and small-world, with a local motif structure similar to that of the TF-TF network (Neph et al.). While interactions are not correlated with the participation in stoichiometric complexes, some RBPs may control each other to regulate cooperative and competitive actions on mutual targets. We identified node chains as the structure of choice for network organization: these widespread regulatory units concur to the formation of a post-transcriptional backbone acting on multiple processes at once to profoundly shape cell phenotypes.

Results

The RBP regulator-of-regulators network

Large-scale mapping of interactions between RBPs and their cognate mRNAs has been conducted by CLIP-like approaches (Milek et al.) in a small number of cellular systems, primarily HEK293, HeLa, and MCF7 cell lines. We previously collected these and other interactions in the AURA 2 database (Dassi, Re et al.). 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 1490 RBPs out of 1793 (of these, 943 are verified and 547 novel as of (Castello et al.); see Experimental Procedures for the RBP list construction details) connected by 19691 interactions, 36 of which (2,41%) are self-loops (i.e. RBPs binding their mRNA). A total of 74 RBPs (4,96%) have outgoing interactions in the network (i.e. they regulate the mRNA of an RBP) mostly coming from CLIP-like assays; the average network degree (number of connections) is 26.4. We also computed the number of individual binding sites for each RBP on each target RBP mRNAs: these reach a maximum of 212 (PCBP2 on the HDGF mRNA), with an average of 5.1 binding sites. 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.) website (http://aura.science.unitn.it).

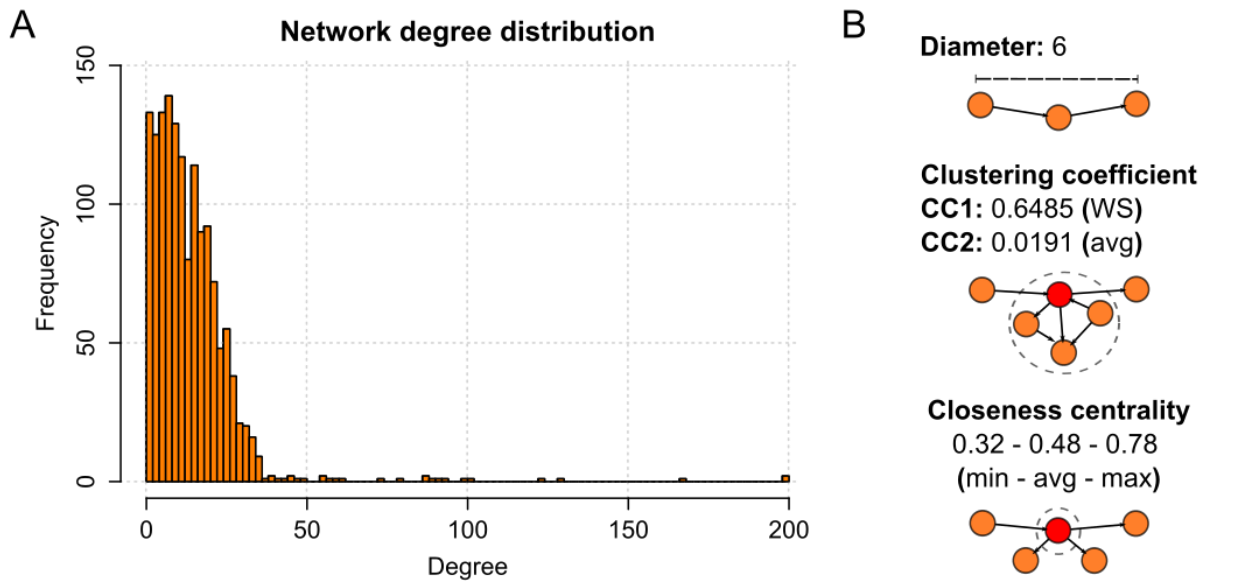

Figure 1: The RBP-RBP network is scale-free and small-world. A) shows the network degree distribution (up to 200), which clearly follows a power-law. **B)** shows the network diameter (top), its average clustering coefficients (middle, Watts-Strogatz 1-neighbor coefficient, named CC1, and 2-neighbor coefficient, named CC2) and closeness centrality (bottom, minimum, average and maximum values for all nodes).

The RBP network has the same global properties of the transcription factors network

We first sought to verify whether the RBP-RBP network has the usual properties of gene regulatory networks, i.e. being "scale-free" and "small-world". To this end, we computed several global and local properties of the network, displayed in **Figure 1**. The degree distribution (Fig. **1A**) is clearly following a power-law, with most nodes having a degree lower than 30 and a minor fraction reaching degrees over 200. This suggests that the network is indeed scale-free, composed of a few central hubs and many peripheral nodes. The diameter (**1B,** D=6) 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=0.6485) which is lost when extending to more distant nodes (2-neighbor coefficient, CC2=0.0191). Eventually, closeness centrality (**1B**, C_c=0.48) reiterates that most nodes are reachable by a small number of steps. We thus quantified the intuitive idea of network small-worldness by computing the S^{WS} measure (Humphries & Gurney), which classifies a network as small-world when greater than 1. We obtained a value of 106.71, 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), 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 & Ruths). We computed the network control profile, which resulted being [s=0.00071, e=0.99929, i=0.0], with *s* representing sources, *e* the external dilations and *i* the internal dilations. Hence, the network is clearly dominated by external dilations (*e*), a fact that locates it in the class of top-down organization systems, aimed at producing a correlated behavior throughout the system: members of this class are transcriptional networks, peer-to-peer systems, and corporate organizations (Ruths & Ruths). These properties also hold in the HeLa and MCF7 networks, suggesting the stability of the network structure in other cellular systems (**Table S2**). We thus focused on the HEK293 network only for subsequent analyses.

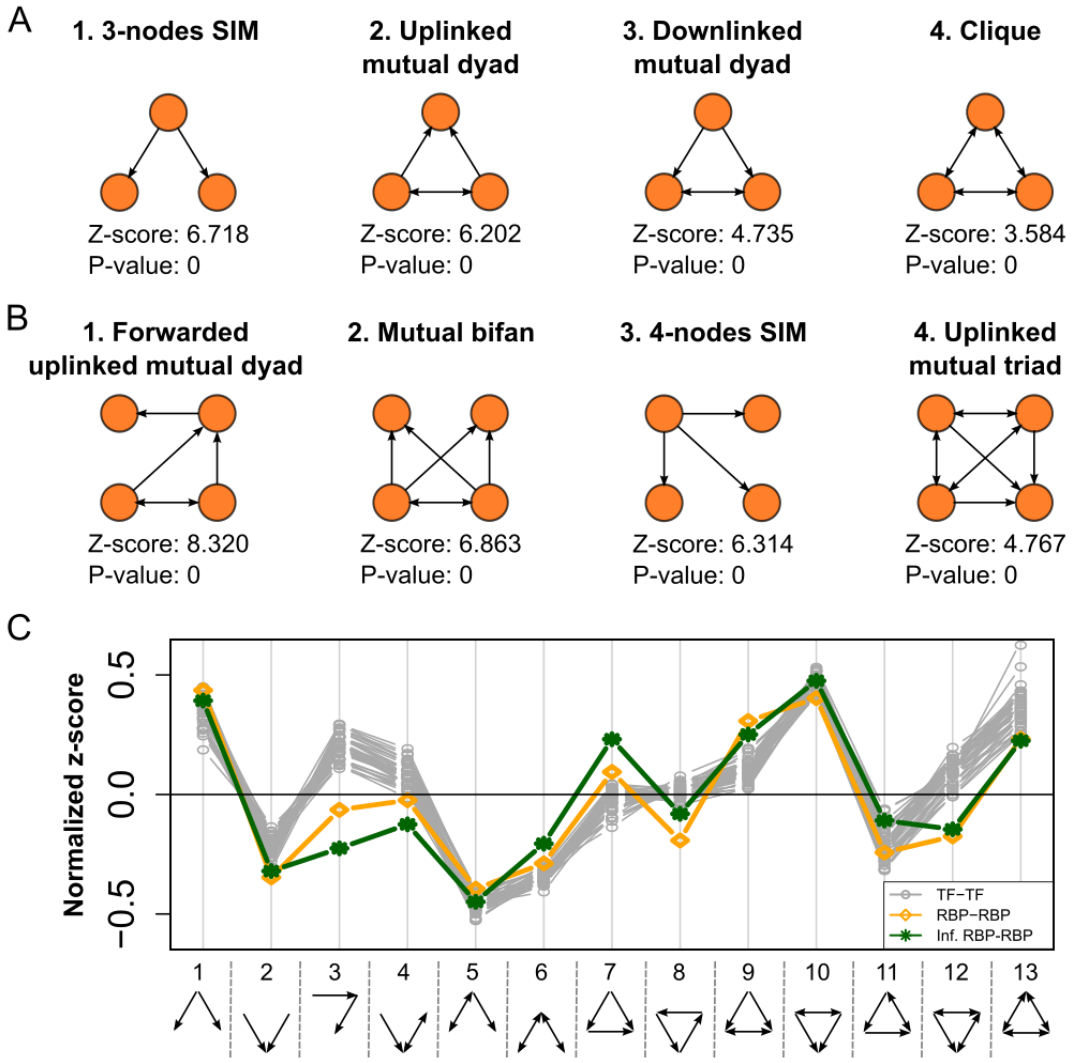

Figure 2: The RBP-RBP network has a distinctive structure. A) shows the most significant 3-nodes motifs identified by FANMOD with their z-score and p-value. **B)** shows the most significant 4-nodes motifs identified by FANMOD with their z-score and p-value. **C)** displays the triad significance profile 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 depletion. While most motifs have similar z-scores in both networks, motifs 3, 4, 7, 8 and 12 are differentially enriched in the RBP-RBP network, suggesting a distinctive network structure.

The RBP-RBP network motif structure is reproducible and distinctive

We then identified motifs (i.e. recurrent patterns of RBP connection) in the local network structure; to this end, we employed FANMOD (Wernicke & Rasche) to look for 3-nodes motifs, of which several patterns have previously been characterized (e.g. feed-forward loop and others (Milo, Shen-Orr et al.)). The most significant motifs are shown in **Figure 2A**: among these the single-input module (SIM) is the most frequent in our network, indicating a widespread use of hub-like patterns. A less frequent motif is the uplinked mutual dyad: RBPs such as *HuR* and *LIN28B* are involved in occurrences of this motif (e.g. both binding to *SYNCRIP* mRNA and each other mRNA). Two other motifs also found at a lower frequency, are the downlinked mutual dyad (realized for example by *TNRC6B*, *HNRNPF,* and *FXR2*) and the fully connected triad, where all three RBPs mutually bind to each other mRNA (e.g. *TDP43*, *PUM2,* and *EWSR1*).

We then identified 4-nodes motifs, the most significant of which are shown in **Figure 2B**: in particular, the 4-nodes SIM extends the 3-nodes SIM, and the uplinked mutual triad extends the alike dyad. The forwarded uplinked mutual dyad motif forwards the output derived from an uplinked mutual dyad to a further RBP, possibly easing the transmission of regulatory information through the network (although likely causing a delay). Eventually, the “mutual bi-fan” belongs to the bi-fans class (Milo et al.), but adds a mutual regulatory link between the two master RBPs, likely to stabilize the regulatory action towards downstream nodes.

To understand whether this network could be considered representative of the unavailable “complete” RBP-RBP network, we built an inferred RBP-RBP network by collecting experimentally determined RBP-bound regions as per a protein occupancy profiling assay in HEK293 cells (Baltz, Munschauer et al.). We then matched these regions to the binding motifs of 193 human RBPs derived by the RNAcompete *in vitro* assays (Ray, Kazan et al.). Only interactions involving two RBPs were included, obtaining a network of 108161 interactions. This

network, independently reconstructed from two experimental datasets, becomes a validation of the general structure we propose for the RBP-RBP network. We thus compared the motif structure of the experimental and inferred RBP-RBP networks with the one of another network of regulators, the TF-TF network described in (Neph et al.) for 41 cell types. We computed the triad significance profile for these networks as described in (Milo, Itzkovitz et al.); the results are displayed in **Figure 2C**. We observe two salient aspects. First, the RBP-RBP network and its inferred version have the same motif structure, with limited magnitude differences only, suggesting that our network structure is reproducible and a representative cross-section of the complete set of RBP-RBP interactions. Then, while the majority of 3-nodes motifs (8/13) have the same enrichment status (normalized z-score resp. greater or smaller than zero) in RBP-RBP and TF-TF networks, five are instead different (enriched instead of depleted or vice versa). This suggests a specialization (although limited) of network structures use for RBP-RBP interactions.

The stoichiometry of RBP complexes does not explain RBP-RBP regulatory interactions

We then sought to hypothesize the biological constraints behind the geometry of RBP-RBP interactions. These might largely be due to interacting RBPs being part of the same complex. To test this hypothesis, we overlapped the interactions in our network with the experimental binary protein-protein interactions (PPIs) contained in STRING (Franceschini, Szklarczyk et al.) and IntAct (Orchard, Ammari et al.). The low amount of network interactions found replicated in PPIs (3.81% for STRING and 0.54% for IntAct) suggests that the network wiring is not made to assure the availability of RBPs for complex assembly. As this analysis dealt with single interactions, we then turned to whole complexes, for some stoichiometric ones (i.e. requiring precise quantities of each of the components for proper functioning) may rely on this mechanism. We employed data from CORUM (Ruepp, Waegele et al.), describing complexes by their composing subunits. A total of 457 interactions overlapped a complex, corresponding to 2.3% of the network only (1000-samples bootstrap p-value <0.001). **Table S3** lists all complexes with at least five interactions in the network involving their subunits. A few well-known complexes are highly represented, including the large Drosha complex (85% of its subunits are in the network, connected by 51 interactions) and the spliceosome (55% of its subunits, connected by 165 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.

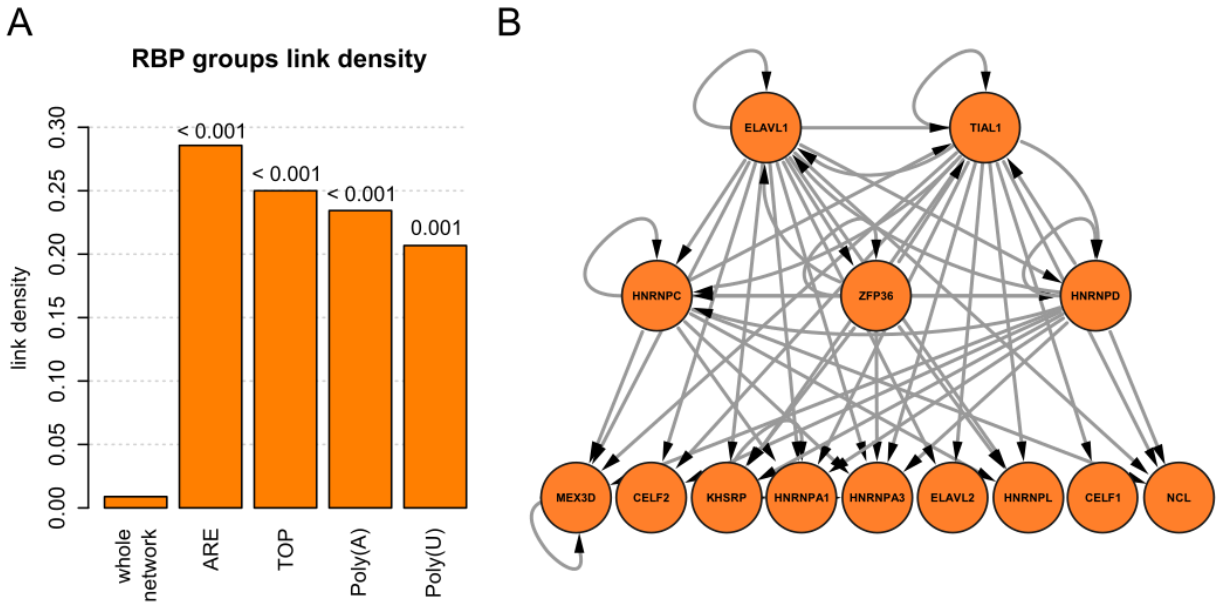

Figure 3: RBPs regulating common sets of targets are more densely interconnected. A) shows link density for the whole network and several groups of target-sharing RBPs: ARE, TOP, Poly(A) and Poly(U)-binding proteins; 1000-samples bootstrap p-values are shown on top of each bar. **B)** shows the complete network of ARE-binding proteins, revealing a hierarchical structure dominated by *ELAVL1* and *TIAL1*.

mRNA target sharing predicts RBP-RBP interactions in the network

RBP-RBP network wiring constraints could be due to combinatorial RBP interactions through their targets. We thus extracted targets for each RBP in the network from AURA 2 (Dassi et al.) and computed the overlap for every RBP pair. We compared these overlaps for protein-mRNA pairs in the network (interacting RBPs) and protein-mRNA not in the network (not-interacting RBPs). The results clearly indicate that interacting RBPs share significantly more targets than non-interacting RBPs (Wilcoxon test $p<2.2E-16$). Looking for the biological meaning of this general phenomenon, we studied groups of RBPs known to bind a well-characterized cis-element. We considered AU-Rich Element (ARE) binding proteins (Barreau, Paillard et al.), proteins interacting with the 5'UTR terminal oligopyrimidine tract (TOP) elements and translationally regulated by the mTORC1 pathway (Hamilton, Stoneley et al., Tcherkezian, Cargnello et al.), proteins interacting with the poly(A), a major cis-determinant of mRNA stability and translation (Goss & Kleiman) and with poly(U) RNAs. ARE-binding proteins, in particular, are known to display both cooperative and competitive behaviors (Barreau et al.). We computed

link density for the whole network and each group: as shown in **Figure 3A**, all groups have significantly higher link densities (22-32 times higher, 1000-samples bootstrap p-values=0.001 or less) than the whole network. The highest density group is the ARE-binding proteins, whose complete network is shown in **Figure 3B**. 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 formed by *ZFP36*, *HNRNPC,* and *HNRNPD,* which then control the remaining RBPs (lowest out-degree). This suggests that such interactions may be needed to regulate cooperative and competitive behaviors on mutual targets.

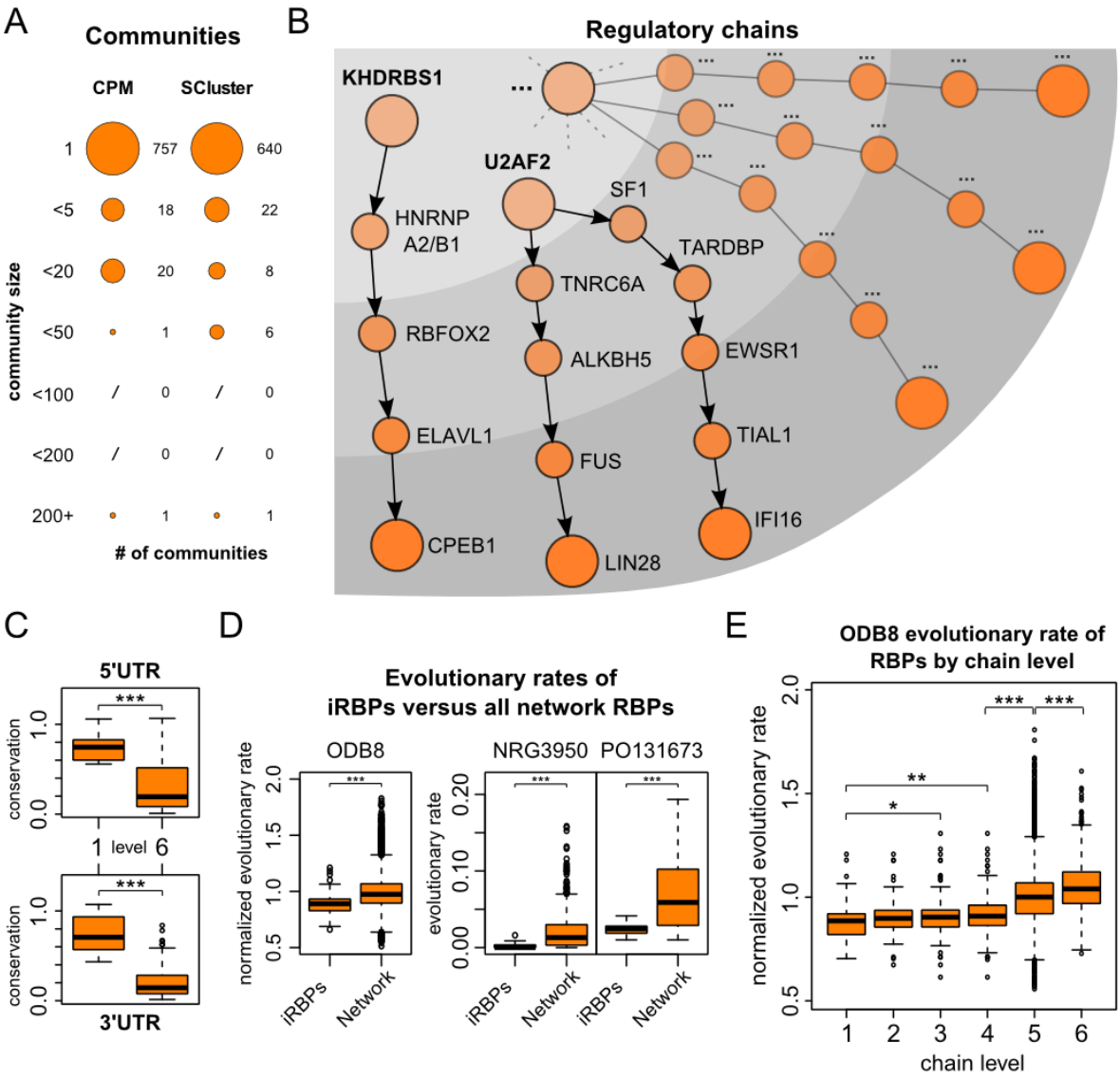

Figure 4: RBP chains are a preferred way of regulatory information transmission in the network. A) shows the low network modularity as per CPM and SCluster communities. Most

are singletons, and one contains almost half of network nodes. **B)** examples of regulatory chains. Dashed lines and dotted name represent an iRBP heading many regulatory chains. Increasing node color intensity represents the "sum" effect of regulatory input through the chain, from the first to the last node**. C)** shows a significant decrease in 5' (top) and 3'UTR conservation (bottom) between the first and the last level of the chains (Wilcoxon test p=8.63E-05 for 5'UTR and 1.16E-06 for 3'UTR). **D)** 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=2.2E-16, 0.0002282 and 0.0004494 for ODB8, NRG3950, and PO131673). **E)** ODB8 evolutionary rates for RBPs along the chains, with a progressive increase which is significant between levels 1-3, 1-4, 4-5 and 5-6 (Wilcoxon test p=0.035, 0.0014, 2.2E-16 and 9.44E-12).

Communities do not determine the structure of the RBP-RBP network

To obtain a more general understanding of RBP-RBP interactions, we then look whether the network has a modular structure, made of RBP communities aimed at regulating specific biological processes. We employed SurpriseMe (Aldecoa & Marin), a tool for the investigation of community structures implementing several algorithms. SurpriseMe is based on Surprise (S) maximization (Aldecoa & Marin), which has been shown to outperform the classic Girvan-Newman modularity measure Q (Newman & Girvan). We used the communities identified by the two best-scoring algorithms implemented in the tool, namely CPM (Palla, Derenyi et al.) and SCluster (Aldecoa & Marin) (S=5614 and 4839). Globally, we observed a low degree of modularity in the network: as shown in **Figure 4A**, 95% of the communities are singletons (i.e. formed by a single RBP) and only a few contain 5 or more RBPs (22 with CPM and 18 with SCluster). Furthermore, both algorithms identified a mega-community including almost one-third of the whole 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. 698 for the RBP-RBP network) and a higher average community size (2.025 for the RBP-RBP network vs. 5 for the TF-TF networks).

A clear association with biological functions does not emerge, if not for a couple of communities. SCluster identifies one where most members are either ribosomal protein or translation factors mRNAs (42 RBPs) while both algorithms find a community composed of the *CPEB* family proteins and several of their targets (24 RBPs for SCluster and 16 for CPM). SCluster and CPM-derived communities are listed in **Table S4A** and **S4B**.

RBP	Chain length	# of reachable RBPs	# of initiated chains	Essential (KBM7)	Essential (HAP1)	Essential (GenomeRNAi)
ALKBH5	4	16	51			v
CPEB1	4	184	2345		v	v
DGCR8	4	24	38	v	v	
EIF3A	4	21	81	v	v	v
EIF3B	4	21	81	v	v	
EIF3D	4	21	81	v	v	
EIF3G	4	21	81		v	v
HNRNPA1	4	128	2301	v	v	v
HNRNPA2B1	4	30	192	v	v	
HNRNPC	4	23	105	v	v	v
HNRNPM	4	164	2085	v	v	v
KHDRBS1	5	31	192			v
	4	317	4573			
KHSRP	4	117	1107			v
PABPC1	4	454	1910	v	v	v
	5	17	21			
PUM1	4	26	70	v	v	
QKI	4	13	15			
RBFOX2	4	19	43			
RBMX	6	128	2423	v	v	
	5	1048	10376			
	4	1393	10839			
RBPMS	4	29	153			
RC3H1	4	161	1764			v
SF1	5	24	62	v	v	

	4	344	4177			
SRSF1	5	127	2423	v	v	v
	4	1049	10402			
STAU1	4	1049	10402		v	
TARBP2	4	16	21			v
TARDBP	4	23	62	v	v	v
TNRC6A	5	17	51		v	
	4	310	3694			
TNRC6C	4	76	752			
U2AF2	5	199	4214	v	v	v
	4	1104	10469			
YTHDF1	4	29	156			
YTHDF2	5	15	27			
	4	455	4036			

Table 1: iRBPs are essential post-transcriptional regulators. The table lists initiator RBPs (iRBPs) for regulatory chains of length 4, 5 and 6. 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, Majek et al.) and in RNAi screenings from GenomeRNAi (Schmidt, Pelz et al.). iRBPs are enriched in essential genes and a significant fraction (50%, bootstrap $p<0.001$) of these is essential in both cell lines. Merging the three sources yields 24/30 iRBPs (80%) as essential.

RBP chains are master regulatory units of the cell

We hypothesized that, instead of communities, the network could be employing node chains as its functional units. To study this aspect, we extracted chains of length 4, 5 and 6 (max. network path length) from the network (examples are shown in **Figure 4B**). We then checked whether these chains were more functionally homogeneous than the communities by computing a semantic similarity score for each pair of RBPs in a chain. We obtained a *functional coherence* score as the average of these similarities, which was also computed for each community composed by more than one RBP. Chains display a significantly higher functional coherence than communities (Wilcoxon test p=6.647E-08/1.053E-10 for CPM/SCluster for chains of length

4; p=3.483E-10/3.835E-13 for CPM/SCluster for length 5; p=6.566E-13/1.659E-15 for CPM/SCluster for length 6; shown as density in **Figure S1**). Chains thus seem a preferred way of functional organization for the RBP-RBP network. In the TF-TF network instead chains are shorter (max. length 5) and 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 minor importance of such units in this network.

Chains are headed by a few initiator RBPs (iRBPs): 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 cell viability. 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.). As shown in **Table 1**, half of the iRBPs are essential in both cell lines, with 19 (63%) essential in the HAP1 line. A 1000-samples bootstrap was significant (p<0.001) in both cell lines and their intersection; furthermore, iRBPs are enriched for essential genes in the same conditions (max. Fisher test p=5.95E-08). A similar amount of iRBPs (16/30) is also essential in at least one cellular model as per RNAi screenings included in the GenomeRNAi database (Schmidt et al.). Merging these annotations with the previously derived ones yields the remarkable total number of 24/30 essential iRBPs (80%). 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**, bootstrap and enrichments are highly significant also for these organisms.

These RBPs could also be highly conserved, due to their fundamental role in driving these regulatory units. We thus computed their average UTR conservation and first found that UTRs of RBPs in the network are more conserved than the UTRs of other genes (Wilcoxon test p=5.489E-10 and <2.2E-16 for 5' and 3' UTRs): as the network includes most RBPs, this feature is characteristic of RBP genes. Then, iRBPs UTR conservation was found to be significantly higher than for other RBPs (Wilcoxon test p=4.19E-07 and 2.01E-06 for 5' and 3'UTRs). Also, as shown in **Fig 4C**, 5' and 3'UTR conservation drops along the chains (conservation is almost stable between nodes 1,2,3 and 4 while it significantly drops at nodes 5 and 6 with Wilcoxon test p=0.0002 and 8.63E-05 for 5'UTR and 9.69E-06 and 1.16E-06 for 3'UTR). We then investigated whether these RBPs are more evolutionarily constrained than other RBPs. We thus obtained evolutionary rates of sequence divergence from the ODB8 database (Kriventseva, Tegenfeldt et al.) and (Zhang & Yang), and rates of purifying selection from (Kryuchkova-Mostacci & Robinson-Rechavi). We observed that, in all datasets, iRBPs have a significantly

lower evolutionary rate with respect to all RBPs in the network (**Fig 4D**; Wilcoxon test p=2.2E-16, 0.0002282 and 0.0004494 for ODB8, NRG3950 and PO131673 resp.). Eventually, we analyzed ODB8 evolutionary rates along the chains, observing a progressive increase (i.e., a progressive decrease in evolutionary constraints) from the first node to the last. This increase is significant between nodes 1-3, 1-4, 4-5 and 5-6 (**Fig 4E;** Wilcoxon test p=0.035, 0.0014, 2.2E-16 and 9.44E-12). These observations, coupled to the essentiality of most iRBPs, consistently support their great importance as key cell regulators.

We then selected *RBMX*, the RBP heading the longest chains and reaching the most other RBPs, for validation. We overexpressed it in HEK293 cells and performed an RNA-seq based transcriptome and translatome profiling to derive the related translational rates (**Figure S2A** shows *RBMX* overexpression), so to get a transcriptome-wide measure of a key functional feature. The results tell us that the regulatory signal sparked by *RBMX* overexpression is both transmitted and amplified along the chains it leads. Indeed, RBPs at the chains final nodes undergo a relevant change in their translational rate, as testified by the translatome fold changes shown in **Figure 5**. The functions of differentially expressed genes we identified after *RBMX* overexpression are shown in **Figure S2B**. Of these 580 genes, 368 (63.4%) are direct targets of the RBPs in the last two levels of *RBMX*-triggered chains. As the regulatory signal is transmitted along the chains, the set of processes controlled by iRBPs may be widely expanded by indirect effects.

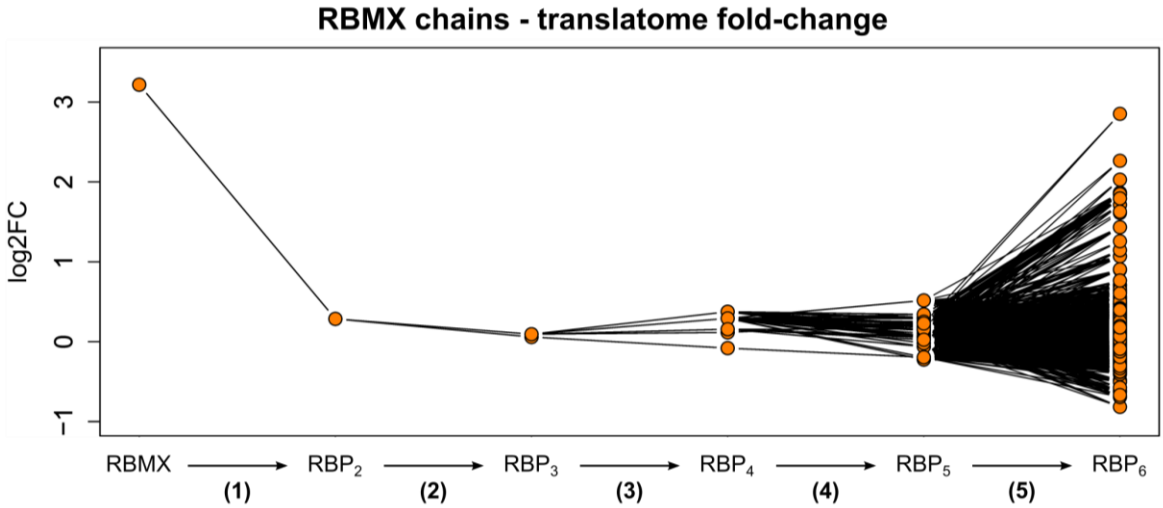

Figure 5: Regulatory information is transmitted along *RBMX*-triggered chains. Displays log2 fold change (at the translatome level) for RBPs at the various levels of *RBMX*-led chains, when over-expressing this gene. The first RBP is *RBMX*, the chain head while RBP_6 represent

the last RBP of a chain (with $RBP_{2..5}$ being the intermediate steps of each chain). Lines represent the RBP-RBP connections in a chain while orange circles represent RBPs.

The RBP-RBP network is a robust and efficient hierarchy

We finally studied the structure of the chains in the network. We measured how hierarchical is the RBP-RBP network (Cheng, Andrews et al.), 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 lower, with a $-\log_{10}p$ of 24.3 versus an average of 3.85 for TF-TF networks at 6 levels. Furthermore, feedback loops (not coherent with a hierarchical organization) are depleted in the network, representing 0.0007% of the motifs only; feed-forward loops, coherent with a hierarchical organization, are instead enriched and amount to 1.26% of the motifs (**Fig. 2C**).

We then assessed the network robustness by computing the pairwise disconnectivity metrics on each node (Potapov, Goemann et al.). The metric is low and significantly lower for the RBP-RBP network than for the TF-TF networks (p-values from 9.1E-73 to 4.98E-90). The network is thus less sensitive to losing a vertex (fewer vertices are disconnected when removing a node), which implies that RBP-RBP interactions are robust, a feature likely granted by the presence of multiple chains.

Eventually, we analyzed redundancy of single chains through the targets of RBPs composing them. We computed the overlap between all targets (both RBPs and non-RBPs) of RBPs at the various levels of each chain. It resulted being particularly low, as only 3.7% of the targets are overlapping between any two levels (median of all chains, average of each pair in a chain; the range is 0.1%-21%). We can thus say that chains are efficient, as targets are not redundantly regulated along the chain nodes, but are predominantly organized in complementary units at each node. The resulting model, coupling hierarchy, network robustness and chain efficiency, is described in **Figure 6**.

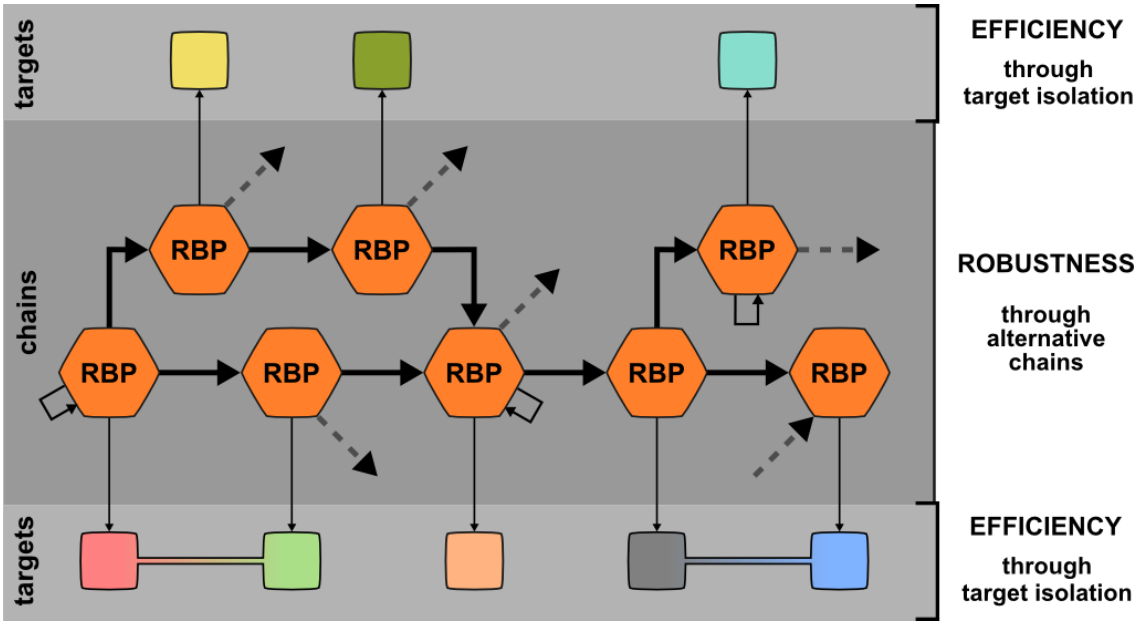

Figure 6: The RBP-RBP network is a robust and efficient hierarchy. Model of the RBP-RBP network as derived from our analyses. RBP-RBP interactions (thick arrows) are robust due to the presence of alternative chains while target mRNA sets (squares) for each RBP in a chain are completely or predominantly different ("isolated") as indicated by the different colors of each square and by the tiny shared area between some mRNA target sets. Dashed in- and out-going arrows hint to the presence of alternative chains while arrows pointing to the originating RBP represent autoregulation events.

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 (Dassi et al., Mukherjee et al., Pullmann et al.), we collected available RBP-mRNA association data and described the network of interactions involving an RBP and an RBP mRNA, also demonstrating that our sampling is likely to represent correctly the whole RBP-RBP network structure.

The network is small-world and scale-free, typical properties of gene regulatory networks. Its local structure is similar to the one of TF-TF networks derived by DNase footprinting (Neph et al.). However, different use of some motifs and the presence of peculiar ones suggest that some degree of structure specialization occurs in the RBP-RBP network with respect to the TF-TF one, likely aimed at better suiting their specific post-transcriptional regulatory task. As said, while the network is partial (as data is available for a fraction of all RBPs only), its motif structure

is highly similar to the one of the inferred network, suggesting that it is representative of general patterns in RBP-RBP interactions.

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 displays this behavior, which cannot thus be considered general. We instead observed that 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.), RBPs can tune the expression of a mutual target by competitive or cooperative binding: we suggest that RBPs may influence the outcome of this process also by regulating the partner RBP. This mechanism could be explained by the need to constrain their expression to yield the intended regulatory effect on mutual targets.

As these behaviors only explain a fraction of the RBP-RBP interactions in the network, we studied instead how RBPs interact with each other. We uncovered the prevalence of chains with respect to communities: modularity measures performed poorly and linear chains are significantly more coherent structures. We believe that RBPs evolved the ability to influence a broad set of biological processes through such chains, which could provide enhanced flexibility with respect to a community pattern. The chains are headed by a few initiator RBPs (iRBPs) expected to heavily shape cell phenotypes. Most iRBPs are essential for the cell, and their UTRs are more conserved than for other RBPs; furthermore, their evolutionary rates are lower, with, remarkably, increasing rates observed along the chain levels. Taken together, these findings truly back iRBP importance as master regulators of key cell processes. Increased levels of one iRBP, RBMX, confirmed the transmission of regulatory information along the chains and the broad scope of affected processes reached by this mechanism. Chains may hence act as “regulatory amplifiers”, allowing substantial responses of the cellular system even with limited initial iRBP perturbations.

Chains profoundly shape the RBP-RBP network to be highly hierarchical and depleted of non-hierarchical loops. Furthermore, we observed robustness at the RBP-RBP interaction level, which is not replicated at the level of RBP-target interaction. Chains are thus efficient, as regulation is not replicated (on average, a target is controlled by only one RBP per chain). While this could in principle lead to a weak architecture, establishing robustness at the RBP-RBP interaction level may be cheaper to obtain, equally effective and potentially more far-reaching.

We suggest that RBP chains use the modulation of RBP targets as a “*connector*” to different processes. Like a piping system bringing the “*fluid*” (regulation) to the various distribution

centers (the RBPs) which then open their "*valve*" (regulate their targets) to influence the functions of interest and pump the "*fluid*" to the next level (next RBP of the chain). Non-chain interactions could then act as chain modulators (e.g. by stopping transmission halfway through, or further enhancing its flow). Under this model, RBP-RBP interactions constitute a post-transcriptional backbone, with RBPs acting as "*split-flow*" pumps to drive regulation and tune protein abundances.

Materials and Methods

Network construction

Regulatory interactions involving two RBPs were extracted from the AURA 2 database (Dassi et al.). The used RBP list included canonical (from InterPro (Hunter, Jones et al.)) and novel RBPs (from (Castello et al.)), for a total of 1793 proteins. 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 were determined by RNA-seq profiles of HEK293 (Kishore, Jaskiewicz et al.), HeLa (Cabili, Trapnell et al.) and MCF7 (Vanderkraats, Hiken et al.), using an expression threshold of 0.1 RPKM. The direction of edges in the network represents regulation by the source RBP on the target RBP mRNA.

The inferred RBP-RBP network was built by collecting RBP-bound regions in mRNA UTRs from a protein occupancy profiling assay in HEK293 cells (Baltz et al.). RNAcompete-derived PWMs for 193 human RBPs (Ray et al.) were downloaded from CISBP-RNA (Ray et al.). Binding regions of these RBP on protein-bound regions were identified by Biopython (Cock, Antao et al.), selecting the best matching RBP for each region (score threshold=0.99); only interactions involving two RBPs (one binding to the other mRNA) were included in the network.

The networks were deposited in NDEX: ID de2d77a0-8480-11e5-b435-06603eb7f303, 5e349554-8481-11e5-b435-06603eb7f303 and c3d85406-8481-11e5-b435-06603eb7f303.

Network properties computation

Network diameter, degree distribution, closeness centrality, Watts-Strogatz (CC1) and two-neighbor (CC2) clustering coefficient were computed by Pajek (Batagelj & Mrvar) and plotted by R (R Core Team). The S^{WS} measure was computed as described in (Humphries & Gurney) by using the Watts-Strogatz clustering coefficient and generating the required random network with Pajek (Batagelj & Mrvar). The network control structure was computed by Zen (Ruths & Ruths).

Hierarchical score was computed as per (Cheng et al.) and pairwise disconnectivity obtained by DiVa (Potapov et al.).

Link density for 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) 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 for motifs of size 3 were computed as described in (Milo et al.) for the RBP-RBP network, the inferred RBP-RBP network and the TF-TF networks described in (Neph et al.).

Communities were studied with the SurpriseMe tool (Aldecoa & Marin): CPM (Palla et al.) and SCluster (Aldecoa & Marin), the algorithms obtaining the highest S values, were used to define communities. Chains were extracted from the network with igraph (Csardi & Nepusz); functional coherence scores were computed with GOSimSem (Yu, Li et al.) by averaging the semantic similarity of each pair of genes in the chain.

Protein-protein interactions and complexes overlap

Human protein-protein interactions were obtained from STRING (Franceschini et al.) and IntAct (Orchard et al.), retaining only interactions of the “binding” type (physical association) and with both partners being in our network. Human protein complexes were downloaded from CORUM (Ruepp et al.). Overlaps were performed by custom Python scripts.

Gene essentiality and phylogenetic conservation analysis

Essential genes of human cells were obtained from (Blomen et al.). Genes associated with an embryonic lethal phenotype in mouse from the MGI (Eppig, Blake et al.); genes associated with a lethality phenotype were extracted from WormBase (Harris, Baran et al.) and FlyBase (dos Santos, Schroeder et al.) for C.*elegans* and D.*melanogaster* respectively. Orthologs of iRBPs were obtained from the same databases. Bootstraps were computed by 1000 random selections of as many genes as iRBPs and computing the fractions of these in the essential genes for each organism.

UTR conservation scores were computed by averaging the phastCons score derived from the UCSC 46-way vertebrate alignment (Karolchik, Barber et al.). 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 (Kriventseva et al.) and two articles (Kryuchkova-Mostacci & Robinson-Rechavi, Zhang & Yang); statistical tests were performed by R (R Core Team).

RBMX plasmid generation

Full-length RBMX was amplified by PCR using HeLa cells cDNA and the following primers: Fw: 5' GAGGCGATCGCCGTTGAAGCAGATCGCCCAGGAA 3' and Rv: 5' GCGACGCGTCTAGTATCTGCTTCTGCCTCCC 3'. The amplified fragment was digested with the SgfI and MluI restriction enzymes and cloned into the pCMV6-AN-His-HA plasmid (PS100017, OriGene, Rockville, MD) to obtain the pCMV6-HIS-HA-RBMX vector, expressing the gene in fusion with an amino-terminal polyhistidine (His) tag and an hemagglutinin (HA) epitope. The construct was confirmed by sequencing.

Cell culture and transfection

HEK293 cells were cultured in DMEM with 10% FBS, 100 U/ml penicillin–streptomycin and 0.01 mM l-glutamine (Gibco, Waltham, MA). Cultures were maintained at 37°C in a 5% CO2 incubator.

1.5×10^6 HEK293 cells were seeded into two 10-cm Petri dishes and transiently transfected using Lipofectamine 2000 (Invitrogen, Waltham, MA) with 2µg of pCMV6-HIS-HA-RBMX or the mock empty vector as control. Total and polysomal RNA extractions were performed 48h post-transfection. All the experiments were run in biological triplicate.

Polysomal fractionation and RNA extraction

Cells were incubated for 4min with 10 µg/ml cycloheximide at 37°C to block translational elongation. Cells were washed with PBS + 10 µg/ml cycloheximide, scraped on the plate with 300 µl lysis buffer (10 mM NaCl, 10 mM MgCl2, 10 mM Tris-HCl, pH 7.5, 1% Triton X-100, 1% sodium deoxycholate, 0.2 U/µl RNase inhibitor [Fermentas Burlington, CA], 10 µg/ml cycloheximide, 5 U/mL Dnase I [New England Biolabs, Hitchin, UK] and 1 mM DTT) and transferred to a tube. Nuclei and cellular debris were removed by centrifugation for 5min at 13,000g at 4°C. The lysate was layered on a linear sucrose gradient (15-50% sucrose (w/v), in 30 mM Tris–HCl at pH 7.5, 100 mM NaCl, 10 mM MgCl2) and centrifuged in a SW41Ti rotor (Beckman Coulter, Indianapolis, IN) at 4°C for 100min at 180,000g. Ultracentrifugation separates polysomes by the sedimentation coefficient of macromolecules: gradients are then

fractionated and mRNAs in active translation, corresponding to polysome-containing fractions, separated from untranslated mRNAs. Fractions of 1 mL volume were collected with continuous monitoring absorbance at 254 nm. Total RNA was obtained by pooling together 20% of each fraction. To extract RNA, polysomal and total fractions were treated with 0.1 mg/ml proteinase K (Euroclone, Italy) for 2h at 37°C. After phenol–chloroform extraction and isopropanol precipitation, RNA was resuspended in 30 µl of RNAse-free water. RNA integrity was assessed by an Agilent Bioanalyzer and RNA was quantified by a Qubit (Life Technologies, Waltham, MA).

Protein extraction and Western blot

10% of each fraction collected from sucrose gradient fractionation was pooled to extract proteins using TCA/acetone precipitation. Proteins were resolved on 15% SDS-PAGE, transferred to nitrocellulose membrane and immunoblotted with HA (Bethyl Laboratories, Montgomery, TX) and RPL26 antibodies (Abcam, Cambridge, UK). Blots were processed by an ECL Prime detection kit (Amersham Biosciences).

RNA-seq

After total and polysomal RNA extraction, 500ng RNA of each sample were used to prepare libraries according to the manufacturer's protocol, using the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA). Sequencing was performed on six lanes at 2x100bp on a HiSeq 2000 machine. Reads were trimmed with trimmomatic (Bolger, Lohse et al.) and aligned to the human hg19 assembly with TopHat2 (Kim, Pertea et al.); gene expression levels and differential expression were assessed with Cufflinks2 (Trapnell, Hendrickson et al.). Functional was performed by Enrichr (Chen, Tan et al.) on Gene Ontology and Reactome annotations. Data were deposited in GEO: ID GSE68990.

Supplemental Figures

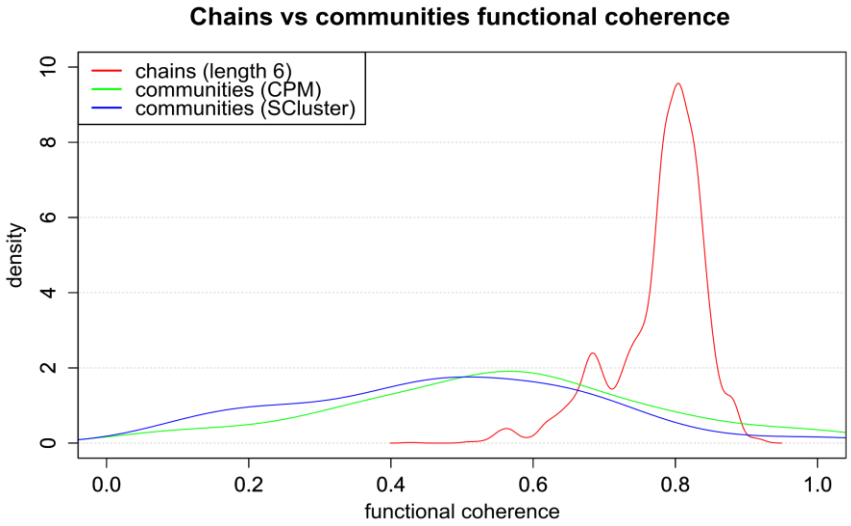

Figure S1: Functional coherence distribution for chains versus communities. Shows the differences in density of the functional coherence values distribution between chains of length 6 (in red) and communities obtained by CPM (green) and SCluster (blue).

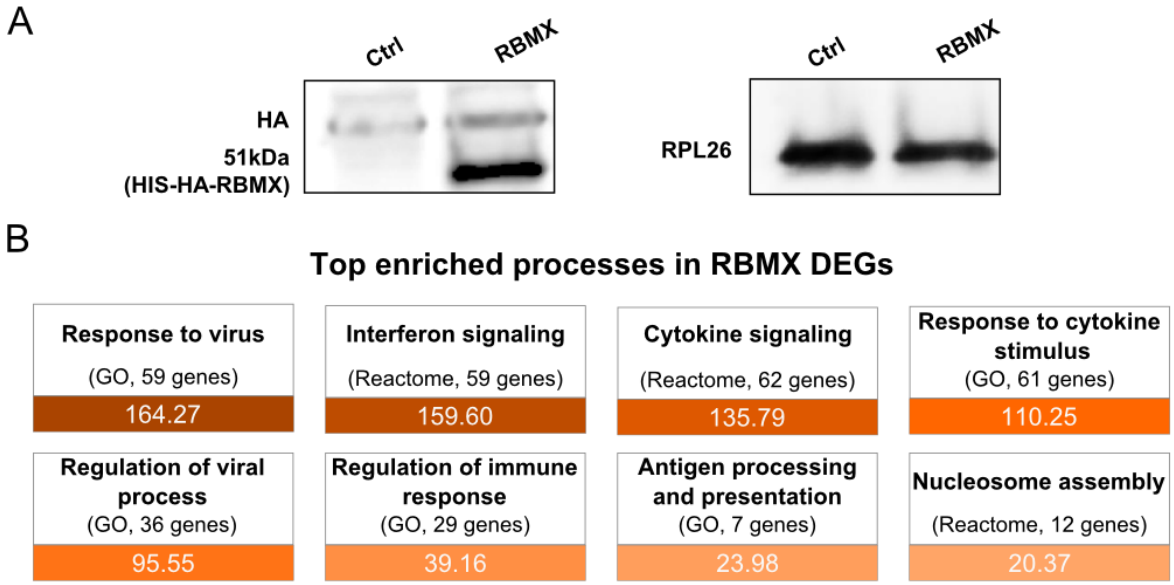

Figure S2: *RBMX* overexpression affects the immune response to viruses. A) *RBMX* western blot in control and overexpressed cells. **B)** top enriched functional themes from Gene Ontology and Reactome, with the associated number of translatome differentially expressed

genes (DEGs) and the functional theme score (computed as log(p-value) * z-score). Color goes from darker (higher score) to brighter (lower score).

Author contributions

ED and AQ designed the research. ED performed data collection and analysis. PZ generated the libraries for RNA-seq. DP and VP produced the RBMX plasmid. ED and AQ wrote the manuscript.

Acknowledgments

We thank Veronica De Sanctis and Roberto Bertorelli (NGS Facility, Centre for Integrative Biology and LaBSSAH, University of Trento) for performing NGS sequencing.

References

Aldecoa R, Marin I (2010) Jerarca: efficient analysis of complex networks using hierarchical clustering. PloS one 5: e11585

Aldecoa R, Marin I (2014) SurpriseMe: an integrated tool for network community structure characterization using Surprise maximization. Bioinformatics (Oxford, England) 30: 1041-2

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell 46: 674-90

Barreau C, Paillard L, Osborne HB (2005) AU-rich elements and associated factors: are there unifying principles? Nucleic acids research 33: 7138-50

Batagelj V, Mrvar A (2002) Pajek— Analysis and Visualization of Large Networks. 2265: 477-478

Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A, Marceau C, Janssen H, Carette JE, Bennett KL, Colinge J, Superti-Furga G, Brummelkamp TR (2015) Gene essentiality and synthetic lethality in haploid human cells. Science

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30: 2114-20
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 & development 25: 1915-27
Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149: 1393-406
Castello A, Fischer B, Hentze MW, Preiss T (2013) RNA-binding proteins in Mendelian disease. Trends in genetics : TIG 29: 318-27
Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics 14: 128
Cheng C, Andrews E, Yan KK, Ung M, Wang D, Gerstein M (2015) An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. Genome biology 16: 63
Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics (Oxford, England) 25: 1422-3
Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Systems 1695: 1695
Dassi E, Re A, Leo S, Tebaldi T, Pasini L, Peroni D, Quattrone A (2014) AURA 2: Empowering discovery of post-transcriptional networks. Translation 2: e27738
Dassi E, Zuccotti P, Leo S, Provenzani A, Assfalg M, D'Onofrio M, Riva P, Quattrone A (2013) Hyper conserved elements in vertebrate mRNA 3'-UTRs reveal a translational network of RNA-binding proteins controlled by HuR. Nucleic acids research 41: 3201-16
dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase C (2015) FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic acids research 43: D690-7
Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database G (2015) The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic acids research 43: D726-36

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic acids research 41: D808-15
Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nature reviews Genetics 15: 829-45
Goss DJ, Kleiman FE (2013) Poly(A) binding proteins: are they all created equal? Wiley interdisciplinary reviews RNA 4: 167-79
Hamilton TL, Stoneley M, Spriggs KA, Bushell M (2006) TOPs and their regulation. Biochemical Society transactions 34: 12-6
Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, Done J, Grove C, Howe K, Kishore R, Lee R, Li Y, Muller HM, Nakamura C, Ozersky P, Paulini M, Raciti D, Schindelman G, Tuli MA et al. (2014) WormBase 2014: new views of curated biology. Nucleic acids research 42: D789-93
Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory network. Nature genetics 39: 683-7
Humphries MD, Gurney K (2008) Network 'small-world-ness': a quantitative method for determining canonical network equivalence. PloS one 3: e0002051
Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D et al. (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic acids research 40: D306-12
Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Babu MM (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. Molecular systems biology 5: 294
Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hinrichs AS, Learned K, Lee BT, Li CH, Raney BJ, Rhead B, Rosenbloom KR, Sloan CA et al. (2014) The UCSC Genome Browser database: 2014 update. Nucleic acids research 42: D764-70
Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nature reviews Genetics 8: 533-43
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 14: R36

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8: 559-64
Kleinberg JM (2000) Navigation in a small world. Nature 406: 845
Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM (2015) OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic acids research 43: D250-6
Kryuchkova-Mostacci N, Robinson-Rechavi M (2015) Tissue-Specific Evolution of Protein Coding Genes in Human and Mouse. PloS one 10: e0131673
Lim C, Allada R (2013) Emerging roles for post-transcriptional regulation in circadian clocks. Nature neuroscience 16: 1544-50
Lukong KE, Chang KW, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends in genetics : TIG 24: 416-25
Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology 8: 479-90
Mansfield KD, Keene JD (2009) The ribonome: a dominant force in co-ordinating gene expression. Biology of the cell / under the auspices of the European Cell Biology Organization 101: 169-81
Milek M, Wyler E, Landthaler M (2012) Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Seminars in cell & developmental biology 23: 206-12
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: 1538-42
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298: 824-7
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. Molecular cell 43: 327-39
Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA (2012) Circuitry and dynamics of human transcription factor regulatory networks. Cell 150: 1274-86
Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Physical review E, Statistical, nonlinear, and soft matter physics 69: 026113

Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L et al. (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic acids research 42: D358-63
Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435: 814-8
Pham H, Ferrari R, Cokus SJ, Kurdistani SK, Pellegrini M (2007) Modeling the regulatory network of histone acetylation in Saccharomyces cerevisiae. Molecular systems biology 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. Molecular and cellular biology 27: 6265-78
R Core Team (2014) R: A Language and Environment for Statistical Computing.
Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale RK, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecenas D, Li W et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. Nature 499: 172-7
Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW (2010) CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic acids research 38: D497-501
Ruths J, Ruths D (2014) Control profiles of complex networks. Science 343: 1373-6
Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M (2013) GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. Nucleic acids research 41: D1021-6
Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature 473: 337-42
Tcherkezian J, Cargnello M, Romeo Y, Huttlin EL, Lavoie G, Gygi SP, Roux PP (2014) Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation. Genes & development 28: 357-71
Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature biotechnology 31: 46-53

Vanderkraats ND, Hiken JF, Decker KF, Edwards JR (2013) Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. Nucleic acids research 41: 6816-27

Vogel C, Abreu Rde S, Ko D, Le SY, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Molecular systems biology 6: 400

Wernicke S, Rasche F (2006) FANMOD: a tool for fast network motif detection. Bioinformatics (Oxford, England) 22: 1152-3

Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics (Oxford, England) 26: 976-8

Zhang J, Yang JR (2015) Determinants of the rate of protein sequence evolution. Nature reviews Genetics 16: 409-20