Interplay between post-transcriptional and post-translational interactions of RNA-binding proteins

Nitish Mittal[1], Tanja Scherrer[2], André P. Gerber[2], Sarath Chandra Janga[3,*]

[1] Biozentrum, University of Basel, Klingelbergstrasse, 50/70 CH-4056, Basel, Switzerland

[2]Institute of Pharmaceutical Sciences, Department of Chemistry and Applied Biosciences, ETH Zurich, CH-8093 Zurich, Switzerland

[3]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL, 61801, USA

[*] *Correspondence should be addressed to :*

Sarath Chandra Janga (sarath@igb.uiuc.edu)

Tel: +1-2172445404, Fax: +1-2173330508

Running title: Regulatory and physical interplay of RNA-binding proteins

Key words: RNA binding proteins, post-transcriptional networks, regulation, network interplay, protein interactions, systems biology

Abbreviations: RBP – RNA binding protein, PTN – Post-transcriptional regulatory network, PPI – Protein-protein interaction, PIN – Protein interaction network, FDR – False Discovery Rate, FFL – Feed-forward loop, FBL – Feed-back loop, RNP – Ribonucleoprotein

Abstract

RNA-binding proteins play important roles in the post-transcriptional control of gene expression. However our understanding of how RBPs interact with each other at different regulatory levels to co-ordinate the RNA metabolism of the cell is rather limited. Here, we construct the post-transcriptional regulatory network among 69 experimentally studied RBPs in yeast, to show that more than one-third of the RBPs auto-regulate their expression at the post-transcriptional level and demonstrate that auto-regulatory RBPs show reduced protein noise with a tendency to encode for hubs in this network. We note that in- and outdegrees in the post-transcriptional RBP-RBP regulatory network exhibit gaussian and scale-free distributions, respectively. This network was also densely interconnected with extensive cross-talk between RBPs belonging to different post-transcriptional steps, regulating varying number of cellular RNA targets. We show that feed-forward loops (FFLs) and superposed FFL/feed-back loops (FBLs) are the most significant 3-node subgraphs in this network. Analysis of the corresponding protein-protein interaction (post-translational) network revealed that it is more modular than the post-transcriptional regulatory network. There is significant overlap between the regulatory and protein-protein interaction networks, with RBPs that potentially control each other at the post-transcriptional level tending to physically interact and being part of the same RNP complex. Our observations put forward a model wherein RBPs could be classified into those which can stably interact with a limited number of protein partners forming stable RNP complexes and others which form transient hubs, having the ability to interact with multiple RBPs forming many RNPs in the cell.

Introduction

Eukaryotic gene expression is a highly complex and regulated process that is controlled at several levels including transcriptional, post-transcriptional and post-translational. Emerging evidence now points to the importance of post-transcriptional control in eukaryotic gene expression [1; 2; 3; 4]. For instance, it has been shown that all post-transcriptional biological properties contribute to 33.15% of the total variation of mRNA-protein correlation [5]. RNA binding proteins (RBPs) play an important role in controlling all the major steps of an mRNA' life, including splicing, export, localization, translation and degradation of mRNA [6; 7; 8; 9]. For instance, Npl3, a yeast SR protein, has been shown to interact with pre-mRNA and regulate all the events from splicing to translational elongation [10]. Similarly, neuronal ELAV protein regulates the fate of its target RNA by mediating the events from poly-adenylation to translation [11]. There are examples of other RBPs which regulate only specific events of mRNA processing, such as Tap protein, like its yeast homolog Mex67, was reported to export mRNA from nucleus to cytoplasm [12]. To facilitate different steps of RNA metabolism, RBPs bind to RNA and form the highly dynamic ribonucleoprotein (RNP) complex. In this complex RBPs associate or dissociate as the RNA metabolism progresses from splicing to translation.

RBPs contain several RNA binding domains that help in binding the target RNA. Some of the most common domains are the RNA recognition motif (RRM), the hnRNP K homology domain (KH) and the Pumilio-Fem3 homology domain (Pum-HD) [2]. In yeast, *saccharomyces cerevisiae*, about 600 RBPs have been predicted on the basis of these RNA binding domains [13]. Other than these putative RBPs (on the basis of previously known RNA binding domains), several metabolic enzymes have also been known to bind to RNA molecules [reviewed in [14]]. For example aconitase (Aco1), a TCA cycle enzyme in yeast *S. cerevisiae* binds to several RNAs encoded by the mitochondrial genome [15]. Similarly, other metabolic enzymes such as glyceraldehyde-3-phosphate dehydrogenase [13] and enolase [16] have also been shown to act as RBPs. These examples indicate the potential for the existence of novel classes of RBPs in eukaryotes. Indeed, two recent studies exploited proteome-wide approaches to identify novel

RBPs in yeast. These studies resulted in the identification of several novel RBPs - most of which were previously characterized as enzymes, suggesting that the dual nature of enzymes is more common than previously thought [17; 18].

RNA-binding proteins provide an additional layer of plasticity in controlling gene expression. They have been shown to be involved in the regulation of several processes such as embryo development in *C. elegans* [19], neuronal differentiation of stem cells [20], T-cell activation [21] etc. To understand the mechanism of how these processes are regulated and affected by RBPs, several large-scale studies have been performed to identify the RNA targets of RBPs [13; 21; 22]. For example, in yeast alone genome-wide studies have identified the targets of several RBPs using RNP immunoprecipitation-microarray (RIP-Chip) method [13; 17; 18]. These studies have revealed that RBPs vary in the number of targets which they regulate depending on their expression level [8], with some RBPs having more than 1000 RNA targets such as Pub1 (1639 targets) and Pab1 (1802 targets) whereas some RBPs such as Nop13, responsible for pre-18s rRNA processing, have as few as 2 RNA targets. These studies also showed that RBPs bind to functionally or cytotopically related targets. For example, Puf3 in yeast binds to cytoplasmic mRNAs of mitochondrial proteins. Likewise, Puf1 and Puf2 have been shown to bind to mRNAs of membrane associated proteins [23]. All these examples support the concept of post-transcriptional operon in eukaryotes [24].

Due to their central role in controlling gene expression at the post-transcriptional level, alteration in expression or mutations in either RBPs or their binding sites in target transcripts have been reported to be the cause of several human diseases such as muscular atrophies, neurological disorder and cancer [reviewed extensively in [25; 26; 27]]. These studies suggest the precise regulation of the expression levels of RBPs in a cell. In fact, a recent systems-wide study of the dynamic properties of yeast RBPs showed that highly connected RBPs are likely to be tightly regulated at the protein level, supporting these observations [8]. Therefore, a central unanswered question in light of these trends is how do RBPs regulate each other at post-transcriptional and post-translational (protein-protein interaction) levels and are there any links between the two levels which govern the interplay between them for proper functioning of RBPs in the context of RNP complexes.

Results

Significant fraction of the RBPs auto-regulate their expression with low protein expression noise and high post-transcriptional connectivity

An important property of several regulatory factors is their ability to regulate their own expression frequently called auto-regulation. Although it is commonly observed for transcription factors in both bacteria and eukarya, there is increasing evidence that RBPs also regulate their expression level [13; 28; 29; 30]. However, it is unclear if auto-regulation of RBPs provides any advantage in controlling their expression level. Therefore, in order to understand whether RBPs which auto-regulate their expression show differences in their expression dynamics compared to those which do not, we have assembled genome-wide RNA targets for a total 69 RBPs in *Saccharomyces cerevisiae* identified by using RIP-CHIP method from different laboratories (see Materials and Methods). Of these, 49 RBPs have been previously screened using this approach by Hogan et.al [13] while an additional 14 and 15 RBPs have been screened in three recent studies from two different groups [17; 18; 31]. Analyzing this dataset for auto-regulatory interactions, we could construct a set of 69 RBPs, of which 26 were found to bind their own RNA (38%) and were termed as auto-regulatory RBPs (ATR) while the rest (43) of the RBPs were termed as non auto-regulatory RBPs (NATR) (see Figure 1 for a complete list). This observation suggests that more than one-third of the RBPs in the cell could be auto-regulating their levels at the post-transcriptional level. To compare the properties which define the expression dynamics of RBPs which regulate their own RNA (i.e, auto-regulatory RBPs) with those which do not regulate their RNA (non auto-regulatory RBPs), several datasets were employed as described in Materials and Methods.

As a result of this analysis (see Supplementary Figure 1) we found that auto-regulatory and non auto-regulatory RBPs do not show any significant difference in their mRNA half-life ($p = 0.97$), mRNA abundance ($p = 0.45$), ribosome occupancy ($p = 0.12$), protein abundance ($p = 0.57$) and protein half-life ($p = 0.91$) by Wilcoxon test, suggesting that RBPs which are auto-regulated do not exhibit any difference in their dynamic regulation compared to those which are not. In contrast to these observations, we found that protein noise of auto-regulatory RBPs is significantly lower than that of non auto-regulatory RBPs ($p < 0.06$). These trends suggest that while auto-regulation of RBPs may not provide any significant advantage in regulating their mRNA or protein turnover rate or even their abundance, it can significantly influence their cell-to-cell variation in protein levels. Noise in protein expression has been associated with a number of phenotypes in various model systems and these results indicate that auto-regulatory RBPs need to be tightly regulated with little variation in their expression levels in a population of cells [32; 33; 34; 35]. Auto-regulation might provide RBPs with an ability to fine tune their expression at post-transcriptional level similar to that observed and proposed as a general phenomenon for transcription factors (TFs) [28; 36]. Indeed, a similar analysis performed for TFs indicates that auto-regulatory TFs in yeast show significant reduction in protein expression noise (Janga et.al, unpublished). One possible explanation for the observed tendency is that auto-regulatory RBPs need an independent control of their expression, as they control a significant fraction of the post-transcriptional network of RBP-RBP interactions (see Materials and Methods and section below). In order to test this hypothesis, we compared the connectivity of auto-regulatory and non auto-regulatory RBPs in this network. This analysis unambiguously revealed that auto-regulatory RBPs are significantly more connected ($p < 2.8E-03$, Wilcoxon test) and tend to have higher indegrees ($p < 0.044$, Wilcoxon test) than non auto-regulatory RBPs in the RBP-RBP post-transcriptional network. These results support the notion that RBPs with high degrees at the post-transcriptional level are likely to be auto-regulatory, as this property would enable them to fine tune their expression in a cellular context. In other words, auto-regulation of highly connected RBPs could possibly control their dosage as well as temporal and spatial concentration within the context of other RBPs enabling them to form appropriate ribonucleoprotein complexes depending on the needs of the cell.

RBP-RBP interactions exhibit a dense network of inherent feedback post-transcriptional links with extensive functional coupling

It is clear from the above analysis that some RBPs bind their own RNA transcripts but others do not. This raises an important question: how are RBPs themselves regulated at the post-transcriptional level? To answer this question we have constructed a network of post-transcriptional interactions between RBPs for all studied RBPs, using currently available RIP-chip data [13; 17; 18; 31]. From the original network which comprised of 69 RBPs and about 25,000 regulatory interactions, we could extract a sub-network where targets included only the studied RBPs (see Materials and Methods). This enabled us to visualize the connections among RBPs due to post-transcriptional control (Figure 1). This directed sub-network comprised of 51 RBPs as regulators and 68 RBPs as targets with a total of 351 interactions among 69 RBPs. The average connectivity in this network is 10.2 (total number of connections between RBPs divided by the total RBPs in the network) with an average out-degree of 6.88 (number of out-going connections) and in-degree of 5.16 (number of in-coming connections) suggesting that RBPs are highly connected to other RBPs through extensive regulatory linkages (Figure 1, Table 1 & Supplementary Table 1). Indeed, we found that the average clustering coefficient of this network was 0.37 (3 times more than that seen in an erdos-renyi random network with the same number of nodes and edges) suggesting a modular organization of this network. This clustering effect is also substantially higher than that observed in an initial study for transcriptional network [37], further supporting strong regulatory connections between RBPs. We also found that while the in-degree was roughly constant for most RBPs, the out-degree showed a scaling distribution indicating that a small set of RBPs might be responsible for controlling a large fraction of them to an equivalent extent (due to their similar in-degrees) (see Supplementary Table 1 and Supplementary Figure 2, correlation for power law fit was 0.87, $y=7.42x^{-0.63}$). The observed indegree distribution is in contrast to that reported for transcriptional networks wherein the fraction of

target genes with a given incoming connectivity was observed to follow an exponential distribution in both *E. coli* and *S. cerevisiae*, while the outgoing connectivity, which is the number of target genes regulated by each transcription factor, was found to be distributed according to a power law, similar to that observed in this study [37; 38]. It is also worth mentioning that RBPs having high out-degree also have more RNA targets but RBPs having very high in-degrees generally tend to have intermediate number of RNA targets (less than 600) (Supplementary Table 1). These observations suggest that RBPs form a dense network of interactions with a small fraction of master regulators controlling a significant fraction of the network.

We found that the average in-degree is high, with up to 5 different RBPs post-transcriptionally controlling a given RBP suggesting that there is extensive cross-talk for using RBPs in a number of different contexts. This was also evident from high betweenness and closeness centralities, which are independent measures for measuring the centrality of a node in complex networks (see Methods, Table 1 and Supplementary Table 2). For instance, we found that the average pathlength to all other nodes (inverse of closeness) from a node of interest is about 2 suggesting that most nodes in this network can be reached within three edges. These observations indicating the dense networking and cross-talk between RBPs was also reflected from an analysis of the post-transcriptional processes these RBPs are associated with (see Ref. [8] for classification) (Supplementary Table 3). For instance, we found that the RBP, Npl3, which promotes elongation, regulates termination, and carries poly(A) mRNA from nucleus to cytoplasm was annotated to be involved in post-transcriptional processes localization, RNA-processing, splicing, translation and transport with more than 1100 RNA targets. Likewise, Pab1 and Nab2 known to control ~2000 and ~700 RNA targets respectively were found to be involved in a number of these processes suggesting that the hubs in this network are also responsible for integrating diverse post-transcriptional events. In other words, those RBPs which have multiple functional labels were found to be hubs and responsible for this cross-talk, as they could enable switching their use depending on the needs of the cell.

We next asked whether the connectivity of a RBP in the RBP-RBP regulatory network has a relation to its dynamic properties such as mRNA or protein turn over, abundance, protein noise etc. Our analysis using a variety of datasets discussed in materials and methods, indicated that noise in protein levels was the best correlated property with RBP degree ($R= -0.375$, $p <1.7E-3$) followed by mRNA half-life ($R = -0.249$, $p <0.039$), protein abundance ($R = 0.245$, $p < 0.042$) and mRNA abundance ($R = 0.219$, $p < 0.07$). We also found that protein half-life and ribosome occupancy did not show any correlation with RBP degree. These trends are generally in line with previous observations made on the complete network of RBP-RNA interactions in yeast [8], except for the mRNA half –life which was found to show a weak positive correlation in the previous study but a weak negative trend in the RBP-RBP network, possibly supporting the notion that hubs in this network might be short-lived as was observed for hubs in the transcriptional regulatory network of *Escherichia coli* [39].

Another hallmark of regulatory networks is the presence of sub-graphs or patterns of interconnections which appear more often than expected by chance and have been referred to as network motifs [30; 40; 41; 42]. In light of these observations, we wanted to understand if the post-transcriptional network of RBP-RBP interactions contains any significant patterns of interconnections. To address this, we employed mfinder- a network motif detection tool, to identify and estimate the significance of different sub-graphs in this network (see Materials and Methods). Our analysis revealed that the most significant three node motifs include i) Feed-forward loops, where in the top-level RBP controls two target transcripts (of RBPs), one of these two RBPs in turn control the second and ii) superposed feed-forward and feed-back loop (FBL) in the same motif, where in all the nodes are RBPs. The later motif comprises of a RBP, X, which controls both the target RBPs Y and Z, with Y controlling Z and Z in turn controlling X. In total, we identified 269 FFL motifs with a p-value < 0.005 and 22 superposed FFL-FBL motifs with a p-value < 0.006 (Supplementary Table 4). Table 2 summarizes the 22 superposed FFL-FBL motifs identified in this network. We found that all of these studied RBPs interact reciprocally to form bidirectional regulatory interactions between the nodes X and Z in this network, suggesting that mutual regulation of RBPs might enable a better co-

ordination in order to guide their post-transcriptional control on a global level. Indeed, we identified a total of 12 unique mutual edges corresponding to 15 different RBPs in this network indicating that mutual regulation might connect different post-translational events and processes to achieve modular functions at this level of regulation. It is important to note that when X and Z have a reciprocal interaction between them, each one can control the other and change the directionality of the FFL resulting in a different affected downstream target. Also note that there is no hierarchy in this motif because X and Z can control each other in a bidirectional fashion, which is in contrast to the FFL motifs where hierarchy is due to the inherent directionality built in them from the top node (X). For instance, in yeast and other eukaryotes non-coding RNA polymerase II transcripts are processed by the poly(A) independent termination pathway which requires a specific factor called the Nrd1 complex [43; 44]. Several studies have shown that this complex comprises of the nuclear pre-mRNA down regulation (Nrd)1 protein, the nuclear polyadenylated RNA-binding (Nab)3 protein as well as the RNA helicase Sen1, interacting with various other RBPs and the exosome, that are needed in the 3' end processing of non-coding RNA transcripts [45; 46; 47; 48]. We found that not only does Nrd1 auto-regulates its own activity but it regulates and is regulated by Nab3 at the post-transcriptional level, in line with recent observations that these RBPs are able to bind to their targets more efficiently when they form a heterodimer to bind in a cooperative manner (i.e, when they are part of the same RNP complex) rather than when they bind as individual subunits [49]. These observations suggest that many of the other RBP pairs detected in Table 2 could be mutually controlling the expression levels of their cognate RBPs at post-transcriptional level in order to define their RNP target space either directly or indirectly through the use an intermediate player represented by Y. It is possible to suggest based on these observations that RBPs can use such mutual post-transcriptional control through direct or indirect loops, to not only decrease the response delay in their expression levels but to generate stoichiometric amounts to control their combined targets under appropriate conditions in the cell.

RBPs forming part of the same ribonucleoprotein complex are likely to have a post-transcriptional regulatory interaction between them

While it is evident from the previous section that RBPs do not work in isolation but are rather controlled by other RBPs at the post-transcriptional level, it is unclear how the post-transcriptional regulatory interactions between RBPs at this level and the network of physical interactions between RBPs are related. So we constructed a network of protein-protein interactions between the experimentally studied RBPs present in the post-transcriptional network, using publicly available datasets (see Materials and methods) to gain an understanding of the underlying principles which allow the integration of these networks. This allowed us to not only study the properties of this protein-protein network but to compare this network with that of the regulatory network in terms of these properties. Protein-protein interactions between RBPs have been known for a long time and most RBPs are known to form dynamic Ribonucleoprotein (RNP) complexes at different stages of their life cycle to perform their functions in the metabolism of RNA [50; 51; 52; 53]. However, currently there is no global analysis of this interactome to compare its properties with that of the corresponding post-transcriptional regulatory network and neither is it clear if these two networks are mutually exclusive or are complementary to each other. To answer the first question, we employed an extensive integrated dataset of protein-protein interactions available for the yeast genome [54] generated by two different groups, using affinity purification protocols [55; 56]. This dataset has more than $4x10^5$ interactions between 5303 proteins in the yeast genome. From this dataset we extracted a subnetwork encompassing the 69 studied RBPs. This subnetwork comprised of a total of 132 interactions between 55 studied RBPs as shown in Figure 2. Analysis of the network properties of this undirected network showed that the average degree of RBPs is about half of that observed for the post-transcriptional network (Table 1). This observation may not be surprising given that the number of edges in the later network is about twice that seen in the protein interaction network of RBPs. Sparse connectivity of the protein interaction network compared to the post-transcriptional network was also evident from the high average betweenness and closeness values observed in the later. However despite the less number of edges, analysis of the average path length and diameter of the networks suggested that protein interaction network

compared to the post-transcriptional network is slightly denser, as is evident from the higher average path length and diameter observed in the later (see Table 1). This was also evident from the high clustering coefficient of the protein interaction network (4 times more than that seen in a erdos-renyi random network with the same number of nodes and edges), which was found to be higher compared to the regulatory network. This increased clustering coefficient suggests a strong modular architecture for the protein interaction network, but surprisingly also for the post-transcriptional network albeit less dramatic. We also found that the protein interaction network had a higher power law exponent value compared to the post-transcriptional counterpart suggesting that physical interactions between RBPs might be exhibiting a scaling behavior in addition to modular organization of the network. Such a scaling and modular organization can be attributed to the organization of multiple protein complexes (modules) in this network, as has been reported for other cellular networks [57; 58].

We next asked if the physical interaction network of RBPs shares any edges with that of the post-transcriptional network to address the extent of overlap between the two levels. This analysis revealed that out of 132 protein-protein interactions, 33 were found to be shared with RBP-RBP regulatory network (Table 3). These 33 interactions represent a significant overlap (p-value < 0.01, hypergeometric) between the two kinds of interaction networks (protein-protein and regulatory interaction). This result suggests that the RBPs which have a regulatory link tend to physically interact with each other to form a RNP complex. This observation also indicates that RBPs which exhibit both physical and regulatory interactions between them are likely to be involved in the same or related post-transcriptional processes to control a common subset of post-transcriptional targets.

Interacting RBPs in the integrated network show co-localization and sharing of targets

Once it is known that RBPs form a dense network both at regulatory as well as protein-protein interaction level, we were interested in the features exhibited by the integrated network formed by the set of interactions detected in both the networks independently. To address this we used the integrated network constructed above. This integrated network comprised of 33 interactions among 32 unique RBPs with differing connectivity's. For instance, Pab1 has maximum connectivity of 9 followed by Bfr1, Puf4 and Gbp2 each with a connectivity of 5 in this integrated network. Most of these RBPs with high connectivity in this network were found to have high number of RNA targets and possibly act as hubs in the regulatory network. While these interactions suggest that there are some RBP pairs that have both a regulatory connection as well as a physical interaction between them; to validate these pairs, we overlapped the protein localization data obtained from a green fluorescent protein (GFP) based study [59] on this network (Table 3). It is interesting to note that a significant proportion of the interacting partners (16 out of 33) co-localized to cellular compartments. For instance, both the interacting partners, Scp160 and Bfr1 in the integrated network localized to the endoplasmic reticulum. An independent study also showed that Scp160 and Bfr1 interact to form a RNP complex to regulate the RNA metabolism and are hence co-localized to endoplasmic reticulum [60]. Similarly, Nrd1 and Nab3 co-localized to nucleus. However there are examples in the integrated network where the interacting partners do not localize to the same compartment. For example, Pub1 resides in the cytoplasm while Nab3 localizes to nucleus. These discrepancies may be due to the fact that the data which has been employed for localization only reported predominant localization of a protein and hence secondary localizations of some of these proteins might not have been reported. For instance, in the case of Pub1, an independent study has shown its localization at both nucleus and cytoplasm [61]. Hence, the co-localization of RBPs can only be considered as one of the evidences to support the presence of protein-protein interaction between RBPs.

In order to further investigate whether the interactions in the integrated network are supported by other means, we asked if two RBPs which interact with each other also share their RNA targets. As we have analyzed only the experimentally studied RBPs for their RNA targets, we were able to investigate for the overlap of RNA targets

between the two interacting RBPs in the integrated network. This analysis strikingly revealed that about 50% (15 out of 33) of the interacting RBPs do show a significant overlap ($p < 0.01$) among their targets (Table 3). This fraction was found to be much higher than when all the 69 RBPs in the entire network were analyzed for the extent of overlap between their targets at the same p-value threshold. In particular, 23.5% of the RBP pairs in the complete network showed statistically significant overlap between their targets, of which 33% of the associations were supported by evidence from either post-transcriptional or post-translational interaction data. P-values in table show the significance for the extent of overlap between all the targets of the two interacting RBPs using hypergeometric distribution. Some RBP pairs such as, Scp160 and Bfr1 (p-value ~ 0), clearly showed a high significance for the overlap of targets as previously reported by Hogan et al. [13]. Scp160 has a total of 1337 targets while Bfr1 binds to 1051 targets and they have 823 targets in common. Few other examples for interacting pairs which share their targets with high significance are Nrd1-Nab3 (p= 7.85E-219), Pub1-Pab1 (p= 1.68E-68) and Pub1-Nrd1 (p= 1.45E-64). However, our analysis also identified interacting pairs which showed low overlap between their target lists. One possible explanation for this lack of overlap is that some RBPs like Pab1 have several interacting partners (9 in this network) in contrast to SGN1, which not only targets fewer RNAs in the cell but also interacts with fewer RBPs, making the former class of RBPs like Pab1 transient players working with many RBPs simultaneously. This is evident from the table with Pab1 showing variation in the overlap of targets with different partners. For instance, Pab1 shows high overlap with Pub1 (p= 1.68E-68), Cbc2 (p= 2.99E-38) and Gis2 (p= 9.03E-34), moderate overlap with Hrb1 (p= 3.8E-03), Mdh3 (p= 0.017) and low overlap with Yra2 (p= 0.17), She2 (p= 0.12), Gus1 (p= 0.22) and Sgn1 (p= 0.15). Another example of this kind is that of Npl3, which was found to show high overlap with Cbc2 (p= 1.43E-59) and weak overlap with Sgn1 (p= 0.15) and Yra2 (p= 0.15). These examples indicate that a single RBP can form different RNP complexes with different RBPs and regulate the different steps of metabolism of diverse groups of RNA at the same time. These observations also point out that even if two RBPs interact with each other (both at the post-transcriptional and post-translational level) they do not necessarily have high degree of overlap between their targets. In contrast to these trends, we also found other class of RBPs - Nab3, Nrd1 and Cbc2 which were found to work in a dedicated manner i.e, they always showed high overlap with all of their interaction partners in the integrated network. These RBPs are likely to be involved in specific processes (such as that discussed in a previous section for Nrd1 complex in poly(A) independent termination pathway) and may be forming stable RNP complexes. RBPs like Pab1 and Npl3 might be involved in a number of post-transcriptional programs by interacting and coordinating with many RBPs to form diverse RNP complexes during their life cycle. In fact, we found that Pab1 and Npl3 were found to physically interact with 28 and 17 other RBPs while Nab3, Nrd1 and Cbc2 had 10, 10 and 6 interaction partners in the RBP-RBP physical interaction network (Supplementary Table2). Further analysis revealed that clustering coefficient can be used as a diagnostic criterion to distinguish these two classes of RBPs, especially when the degree of the RBPs being compared is high. For instance, we found that in general transient RBPs like Pab1 and Npl3 had high degree but low clustering coefficient (0.13 and 0.33 respectively) in the protein-protein interaction network while stable RBPs like Nab3, Nrd1 and Cbc2 were found to have very high clustering coefficient (0.77, 0.77 and 0.93, respectively) despite having moderate number of interaction partners (Supplementary Table 2). These observations suggest that transient RBPs involved in multiple post-transcriptional processes can be distinguished from the stable ones based on their clustering coefficient in the protein interaction network.

Discussion

In this study, we have developed an integrated map of RNA-protein (post-transcriptional network of RBPs) and of RBP-RBP (post-translational interaction network of RBPs) interactions to get a first comparative understanding of the interplay between the two levels. For instance, in yeast, analysis of the transcriptional regulatory network of transcription factors showed the presence of feedback loops i.e. the target transcription factor controls the expression

of its regulator [29]. Therefore, we wanted to evaluate if similar properties are also obeyed in the post-transcriptional regulatory network of RBPs in yeast, *Saccharomyces cerevisiae*. At the post-translational interaction (physical interactions between RBPs) level, RBPs are known to work in the form of RNPs - a dynamic protein complex where dissociation and association of RBPs takes place as the RNA metabolism progresses. Analysis at this level allowed us to address whether RBPs regulated by other RBPs at the post-transcriptional level can become part of the same ribonucleoprotein complex. This question helped us to understand the reason for the regulatory link between RBPs - is it because the RBPs, which have regulatory link, work together as part of same RNP complex to regulate gene expression or are the two levels mutually exclusive in their control? Finally, analysis of the integrated network constructed using the overlapping set of protein-protein and regulatory interactions between the studied RBPs showed that most pairs of interacting RBPs not only share their localization but also their RNA target pool.

During the last decade enormous efforts have been put to understand the transcriptional regulatory mechanisms in several model organisms, however our understanding of them at the post-transcriptional level is rather limited. Recent high-throughput studies have enabled us to start addressing genomic principles governing the post-transcriptional control by RBPs in these model systems. In this study, we have attempted to dissect this layer of the regulatory network by constructing a network of post-transcriptional interactions between RBPs in yeast. This enabled us to show that a striking fraction of the RBPs in yeast autoregulate their own expression. A deeper analysis of the dynamic properties which can explain this observed trend indicated that auto-regulatory RBPs exhibit low protein expression noise and are usually highly connected in the network of RBPs, suggesting that auto-regulation of these RBPs might provide a means of independent control of their expression thereby providing quick and timely response to intracellular changes or due to external perturbations. This observation has important implications in driving the gene expression changes during development or in cellular differentiation, where a number of RBPs are known to be actively involved in controlling the fate of the transcripts in a just-in-time fashion [62]. An analysis of the RBP-RBP regulatory network showed that it forms a dense intertwined network of post-transcriptional feedback loops with extensive crosstalk between RBPs belonging to different steps of RNA metabolism. A closer examination of the directed network of post-transcriptional interactions between RBPs indicated that the indegree of RBPs in this network followed a gaussian distribution while outdegree followed a scale-free distribution. We also found that FFL which is known to be the most common motif in the transcriptional networks is also the most prevalent structure in the RBP-RBP post-transcriptional network. In addition to the known FFL motif, we discovered the existence of a superposed FFL and FBL motif which composes of a mixed feed-back and feed-forward mechanism co-existing in the same motif as yet another significant pattern in this network, providing evidence for the extensive cross-talk between RBPs to control related post-transcriptional events and processes.

A similar analysis of the protein-protein interactions between the same set of studied RBPs showed that the corresponding post-transcriptional network is less modular, suggesting the existence of RNP complexes inherent in the protein interaction network. We also found that RBPs in the same RNP complex are likely to regulate each other at the post-transcriptional level. This strong trend possibly supports the notion that most RBPs may use their regulatory component in order to dissociate and associate to form dynamic RNP complexes depending on the growth conditions and needs of the cell as the cell cycle progresses. Our observation that RBPs which interact at post-translational (protein interaction) as well as at the post-transcriptional level often show co-localization and significant sharing of targets, suggests that there are two classes of RBPs i) those which stably interact with a limited number of partners so that they share significant number of targets with the interacting partners and ii) those which are transient and interact with multiple RBPs forming many RNPs during their life cycle aiding multiple steps in the metabolism of RNA. Such transient RBPs could act as linkers between different post-transcriptional steps/modules and could mediate their appropriate usage depending on the requirements of the cell. Whether such a classification would be generic to other eukaryotes or even to other yeasts is an open question which can only be addressed as more high-throughput data becomes available for other model organisms.

In summary, our results show that RNA-binding proteins not only form densely interacting networks at both the post-transcriptional and post-translational levels, but also integrate these two distinct levels for coordinating their cellular roles. Our analysis also supports that a significant fraction of the RBPs regulate their expression, either directly by autoregulation or indirectly via other RBPs, at post-transcriptional level.

Materials and methods

Construction and analysis of post-transcriptional regulatory network among RBPs

To understand the cross-talk at the post-transcriptional level between RBPs, we have first constructed the posttranscriptional regulatory network which is a directional network, with RBPs as regulators and RNA molecules of all the experimentally studied RBPs as targets. To construct this network we have integrated four recently published datasets reporting the RNA targets of RBPs in yeast generated by immunoprecipitation of RBPs followed by microarray analysis of the bound transcripts [13; 17; 18; 31]. Briefly, these downloaded datasets comprise of i) 49 previously studied RBPs and their RNA targets from Hogan et al. [13] comprising about 16924 interactions, ii) the post-transcriptional network for 13 unconventional RBPs reported by us earlier [18], iii) a network of 7636 post-transcriptional links for 15 unconventional RBPs reported by Tsvetanova et. al [17] and iv) a focused study on Gis2 reporting more than 700 RNA targets [31]. From these datasets, we have excluded those RBPs for which the original studies have reported any potential interference of the TAP-tag with RNA-binding, those RBPs which had unusually low number of targets and/or show binding to non-protein coding regions or the RBPs have not been reported as bonafide RBPs (as is the case with Smy1 and Mtq2 in the Tsvetanova study). After filtering the datasets at False Discovery Rates (FDRs) and p-values reported to be stringent thresholds for obtaining high-quality interactions in the original studies, removing any redundant interactions between datasets and excluding features which do not encode for protein coding regions, we integrated the resulting data to generate a large compendium of RBP-RNA interactions for yeast. This compilation has enabled us to construct a network of 69 RBPs and 24,932 RBP-RNA interactions on a genome-wide scale. From this network, we have extracted a sub-network, where targets included only the studied RBPs, to understand the regulation of RBPs by other RBPs at the post-transcriptional level. This final sub-network among 69 RBPs comprised of 51 RBPs as regulators and 68 RBPs as targets, with a total of 351 interactions, allowing us to analyze different properties of the network. Auto-regulatory RBPs in this network were defined as those RBPs which bind their own transcript to control their expression at post-transcriptional level. We identified 26 out of the 69 RBPs in this study to be auto-regulated.

Calculation of network properties

To study the properties of the RBP-RBP network and to understand the centrality of the nodes in this framework, we used igraph, a publicly available R package for analyzing graphs [see http://cneurocvs.rmki.kfki.hu/igraph/ and http://www.r-project.org]. In particular, since the network of post-transcriptional interactions analyzed in this study is directed, we used the corresponding versions of the functions: degree, transitivity, betweenness and closeness for calculating the degree (connectivity), clustering coefficient, betweenness and closeness centralities of a node. It is important to note that since the network is directed each node can have both an out-degree, which defines the number of outgoing connections, as well as in-degree, which relates to the number of incoming interactions. Betweenness centrality, which is the number of shortest paths going through a node was calculated using the brandes algorithm [63] implemented in R. Similarly, closeness, measured as average length of the shortest paths to all the other vertices in the graph, was obtained using the implementation in R. Since the centrality measures,

betweenness and closeness use the shortest path lengths between all pairs of nodes in a graph, for cases where no path exists between a particular pair of nodes, shortest path length was taken as one less than the maximum number of nodes in the graph. Note that this is also the default assumption for calculating centrality measures in igraph. Clustering coefficient is a property of a node which tells how connected are the neighbors of a given node to what is expected when all the neighbors are completely connected. An extension of this metric to the complete network defined as the average clustering coefficient tells whether the network is modular or is sparsely connected. Network properties for the protein-protein interaction network of RBPs were calculated using the corresponding undirected versions of the functions in igraph where appropriate. Note that the integrated network was considered directionless and hence degree refers to the total number of connections of a node in this case. To evaluate whether the degree distribution of a network follows a power law or a scale-free distribution, we used the power.law.fit function available in igraph which provides an estimate of the exponent alpha to define the likelihood of a good power law fit. To compare the extent of clustering coefficient observed in the RBP-RBP post-transcriptional and protein-protein interaction networks to a null model, we constructed random networks of the erdos-renyi type by maintaining the number of nodes and edges. Briefly, in an erdos-renyi random network, each node has equal probability to be connected to other nodes in the network, independent of the other nodes.

Data for comparative analysis of expression dynamics

To study the expression dynamics of RBPs in comparison to other groups of genes, we have employed a variety of datasets. These include the transcript stability [64], mRNA copy number, ribosome occupancy [65], protein half-life [66], protein abundance [67] and protein noise [68]. Transcript stability which is measured as the RNA half-life of a transcript could be obtained for 4687 genes in the entire genome while the translational rate defined by the ribosome occupancy and the number of mRNA copies of a gene described by the parameter mRNA copy number per cell could be obtained for 5700 and 5643 genes respectively, allowing us to study the translation rates of the transcripts and the extent of transcript abundance. In yeast, protein half-lives have been estimated by Belle and co-workers for about 3750 proteins by inhibiting translation [66]. Protein abundance which reveals the absolute number of protein molecules per cell was obtained from Ghaemmaghami et. al [67]. We could obtain abundance values for 3868 proteins in the entire genome. Biological noise which is typically defined as the variation in the expression of a protein between different cells in a homogenous population of cells was obtained from Newman et. al [68]. We could obtain noise data for 2213 genes for cells grown on rich media.

Calculation of statistical significance

To assess whether RBPs exhibit a different trend compared to non-RBPs for each of the dynamic properties studied, we used the Wilcoxon rank-sum test or the Mann–Whitney U test available in the R statistical package to calculate the significance. The Wilcoxon test enables the comparison of two samples to assess whether they come from the same distribution or not. Since this test is nonparametric and does not assume any inherent distribution of the samples it is ideal to compare different samples of similar or dissimilar sizes. For studying the overlap between protein-protein and post-transcriptional interactions, we used hypergeometric probability function available in R. Other statistical tests are used as appropriate throughout the text and are cited at relevant places.

Identification of network motifs

Network motifs are defined as recurring regulation patterns which occur in the networks more often than expected by chance [36; 41; 42]. In the regulatory network of *E. coli* and other organisms three distinct types of motifs have been found to be predominant, namely (i) FFL, in which a TF regulates the expression of another transcription factor which together modulate the expression of the target gene; (ii) SIM, in which a single TF regulates several genes

and is equivalent to a simple regulon [69]; (iii) DOR, in which different TFs regulate overlapping sets of genes and are analogous to complex regulons [69]. FFL appears to be the most abundant motif among the best studied transcriptional networks. To identify different kinds of motifs in the post-transcriptional regulatory network of RBPs we searched for the sub-graphs of different sizes in the network using the motif finding tool, mfinder [41]. In order to calculate the significance of an observed sub-graph, we generated 1000 randomly generated networks with the same topology as the original RBP-RBP network in mfinder, which allowed us to compare the occurrences and obtain a Z-score and p-value. Our analysis resulted in the identification of only 2 three-node motifs as significant. These included a total of 269 FFLs with a Z-score of 2.47 and p-value < 0.005 and 22 superposed FFL-FBL motifs with a Z-score of 2.73 and p-value < 0.006 in the network. Complete set of these network motifs belonging to these two categories identified in this study can be obtained as Supplementary Table 4.

Construction of protein-protein interaction network for RBPs

To understand the protein-protein interaction network of the studied RBPs, we obtained a comprehensive map of protein-protein interactions available for the yeast genome by integrating two different high-throughput studies [54]. Briefly, in this study the integration of the mass spectrometry data involved the calculation of purification enrichment scores for each of the interactions in the two datasets [55; 56] and then calculating a combined confidence score for the interactions, in order to account for replicating interactions between the datasets. It is noteworthy to mention that this is the largest available co-complex interactome data available for yeast and hence is also likely to have high confidence indirect interactions i.e, as long as they belong to the same complex, which makes this dataset ideal for this study. We have employed the confidence scores as a metric to obtain the final set of interactions for the analysis in this study. This dataset comprised of 5303 proteins and 401821 interactions. From this network, a sub-network of protein interactions among the 69 studied RBPs was obtained. This resulted in a network of 132 protein-protein interactions between 55 RBPs in the original network. Analysis of this sub-network for network properties was performed using the igraph package as described above. Further, this protein-protein interaction network of studied RBPs was compared with the regulatory network of RBPs to extract only those protein-protein interactions which also have a regulatory connection among them.

Acknowledgements

This work was supported by UNIBAS to NM, a Fellowship from Institute for Genomic Biology at UIUC to SCJ. This work was supported by grants from the Swiss National Science Foundation (3100A0-112235) and the Bonizzi-Theler Foundation to A.P.G.

Supplementary information:

Supplementary figures, tables supporting this work and the network of RBP-RNA interactions used in this study can be obtained from the webpage - http://www.mrc-lmb.cam.ac.uk/genomes/sarath/RBP-RBP-interactome

References

1. Mata, J., Marguerat, S. & Bahler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **30**, 506-14.
2. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**, 1977-86.
3. Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-43.

4. Halbeisen, R. E., Galgano, A., Scherrer, T. & Gerber, A. P. (2008). Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell Mol Life Sci* **65**, 798-813.
5. Wu, G., Nie, L. & Zhang, W. (2008). Integrative analyses of posttranscriptional regulation in the yeast Saccharomyces cerevisiae using transcriptomic and proteomic data. *Curr Microbiol* **57**, 18-22.
6. Sanford, J. R., Gray, N. K., Beckmann, K. & Caceres, J. F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* **18**, 755-68.
7. Zhong, X. Y., Wang, P., Han, J., Rosenfeld, M. G. & Fu, X. D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* **35**, 1-10.
8. Mittal, N., Roy, N., Babu, M. M. & Janga, S. C. (2009). Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proc Natl Acad Sci U S A* **106**, 20300-5.
9. Sette, C. (2010). Post-translational regulation of star proteins and effects on their biological functions. *Adv Exp Med Biol* **693**, 54-66.
10. Gross, T., Richert, K., Mierke, C., Lutzelberger, M. & Kaufer, N. F. (1998). Identification and characterization of srp1, a gene of fission yeast encoding a RNA binding domain and a RS domain typical of SR splicing factors. *Nucleic Acids Res* **26**, 505-11.
11. Pascale, A., Amadio, M. & Quattrone, A. (2008). Defining a neuron: neuronal ELAV proteins. *Cell Mol Life Sci* **65**, 128-40.
12. Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B. K. & Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell* **1**, 649-59.
13. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. (2008). Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6**, e255.
14. Ciesla, J. (2006). Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim Pol* **53**, 11-32.
15. Chen, X. J., Wang, X., Kaufman, B. A. & Butow, R. A. (2005). Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science* **307**, 714-7.
16. Entelis, N., Brandina, I., Kamenski, P., Krasheninnikov, I. A., Martin, R. P. & Tarassov, I. (2006). A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in Saccharomyces cerevisiae. *Genes Dev* **20**, 1609-20.
17. Tsvetanova, N. G., Klass, D. M., Salzman, J. & Brown, P. O. (2010). Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PLoS One* **5**.
18. Scherrer, T., Mittal, N., Janga, S. C. & Gerber, A. P. (2010). A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* **5**, e15499.
19. Lee, M. H. & Schedl, T. (2006). RNA-binding proteins. *WormBook*, 1-13.
20. Ratti, A., Fallini, C., Cova, L., Fantozzi, R., Calzarossa, C., Zennaro, E., Pascale, A., Quattrone, A. & Silani, V. (2006). A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of Msi1 mRNA. *J Cell Sci* **119**, 1442-52.
21. Mukherjee, N., Lager, P. J., Friedersdorf, M. B., Thompson, M. A. & Keene, J. D. (2009). Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol Syst Biol* **5**, 288.

22. Amorim, M. J., Cotobal, C., Duncan, C. & Mata, J. (2010). Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Mol Syst Biol* **6**, 380.
23. Gerber, A. P., Herschlag, D. & Brown, P. O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* **2**, E79.
24. Keene, J. D. & Tenenbaum, S. A. (2002). Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* **9**, 1161-7.
25. Lukong, K. E., Chang, K. W., Khandjian, E. W. & Richard, S. (2008). RNA-binding proteins in human genetic disease. *Trends Genet* **24**, 416-25.
26. Musunuru, K. (2003). Cell-specific RNA-binding proteins in human disease. *Trends Cardiovasc Med* **13**, 188-95.
27. Kim, M. Y., Hur, J. & Jeong, S. (2009). Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep* **42**, 125-30.
28. Martinez-Antonio, A., Janga, S. C. & Thieffry, D. (2008). Functional organisation of Escherichia coli transcriptional regulatory network. *J Mol Biol* **381**, 238-47.
29. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. & Young, R. A. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804.
30. Kanitz, A. & Gerber, A. P. (2010). Circuitry of mRNA regulation. *Wiley Interdiscip Rev Syst Biol Med* **2**, 245-51.
31. Scherrer, T., Femmer, C., Schiess, R., Aebersold, R. & Gerber, A. P. (2011). Defining potentially conserved RNA regulons of homologous zinc-finger RNA-binding proteins. *Genome Biol* **12**, R3.
32. Raj, A. & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-26.
33. Maamar, H., Raj, A. & Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317**, 526-9.
34. Raser, J. M. & O'Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010-3.
35. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451-64.
36. Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-61.
37. Guelzim, N., Bottani, S., Bourgine, P. & Kepes, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* **31**, 60-3.
38. Thieffry, D., Huerta, A. M., Perez-Rueda, E. & Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays* **20**, 433-40.
39. Wang, E. & Purisima, E. (2005). Network motifs are enriched with transcription factors whose transcripts have short half-lives. *Trends Genet* **21**, 492-5.
40. Janga, S. C. & Collado-Vides, J. (2007). Structure and evolution of gene regulatory networks in microbial genomes. *Res Microbiol* **158**, 787-94.

41. 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.
42. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* **31**, 64-8.
43. Carroll, K. L., Ghirlando, R., Ames, J. M. & Corden, J. L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* **13**, 361-73.
44. Carroll, K. L., Pradhan, D. A., Granek, J. A., Clarke, N. D. & Corden, J. L. (2004). Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol* **24**, 6241-52.
45. Steinmetz, E. J. & Brow, D. A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol* **16**, 6993-7003.
46. Steinmetz, E. J., Conrad, N. K., Brow, D. A. & Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**, 327-31.
47. Vasiljeva, L. & Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* **21**, 239-48.
48. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* **91**, 457-66.
49. Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S. & Stefl, R. (2010). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (Nab)3 protein. *J Biol Chem*.
50. Luna, R., Gaillard, H., Gonzalez-Aguilera, C. & Aguilera, A. (2008). Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* **117**, 319-31.
51. Anderson, P. & Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat Rev Mol Cell Biol* **10**, 430-6.
52. Sephton, C. F., Cenik, C., Kucukural, A., Dammer, E. B., Cenik, B., Han, Y. H., Dewey, C. M., Roth, F. P., Herz, J., Peng, J., Moore, M. J. & Yu, G. (2010). Identification of neuronal RNA targets of TDP-43-containing Ribonucleoprotein complexes. *J Biol Chem*.
53. Collins, L. J., Kurland, C. G., Biggs, P. & Penny, D. (2009). The modern RNP world of eukaryotes. *J Hered* **100**, 597-604.
54. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S. & Krogan, N. J. (2007). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol Cell Proteomics* **6**, 439-50.
55. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A. & Greenblatt, J. F.

(2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-43.
56. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B. & Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-6.
57. Ravasz, E. & Barabasi, A. L. (2003). Hierarchical organization in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **67**, 026112.
58. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-5.
59. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**, 686-91.
60. Lang, B. D., Li, A., Black-Brewster, H. D. & Fridovich-Keil, J. L. (2001). The brefeldin A resistance protein Bfr1p is a component of polyribosome-associated mRNP complexes in yeast. *Nucleic Acids Res* **29**, 2567-74.
61. Anderson, J. T., Paddy, M. R. & Swanson, M. S. (1993). PUB1 is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in Saccharomyces cerevisiae. *Mol Cell Biol* **13**, 6102-13.
62. Thomsen, S., Anders, S., Janga, S. C., Huber, W. & Alonso, C. R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol* **11**, R93.
63. Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* **25**, 163-177.
64. Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D. & Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**, 5860-5.
65. Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O. & Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 3889-94.
66. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* **103**, 13004-9.
67. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature* **425**, 737-41.
68. Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. & Weissman, J. S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-6.
69. Gutierrez-Rios, R. M., Rosenblueth, D. A., Loza, J. A., Huerta, A. M., Glasner, J. D., Blattner, F. R. & Collado-Vides, J. (2003). Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. *Genome Res* **13**, 2435-43.

Figure Legends

Figure 1: Network of post-transcriptional regulatory interactions between RBPs. Each node corresponds to one RBP with the links between them corresponding to post-transcriptional regulatory interactions. The network is laid out using circular organization in cytoscape and reflects the dense set of interconnections between various RBPs. RBPs which bind to their own mRNA and hence auto-regulate their transcript level are shown with loops in red color. All other interactions are shown in blue.

Figure 2: Network of protein-protein interactions among the studied RBPs. Each node corresponds to one RBP and the physical interaction between the RBPs is shown as an edge. The network is laid out using circular layout in cytoscape.

Tables

Table 1. Properties defining the structure of the post-transcriptional network (PTN) and physical interaction network (PIN) among the studied RBPs in *S. cerevisiae*. All the network properties are calculated using igraph, a publicly available R package for analyzing graphs (see Methods). Average values are shown for properties which can be calculated for each node in the network.

Network property	Definition	PTN	PIN
Degree or Connectivity	Degree or connectivity refers to the number of interactions a RBP has in this network – the higher the connectivity (*i.e.*, hub nodes) the more the number of RNA targets or more the number of physically interacting RBPs the node has.	10.2	4.8
Indegree	Number of incoming connections to a RBP in a directed network (PTN in this case). Undirected networks do not have this property.	5.16	NA
Outdegree	Number of outgoing connections to a RBP in a directed network (PTN in this case). Undirected networks do not have this property.	6.88	NA
Clustering coefficient	Clustering coefficient of a RBP reflects the extent to which the neighbors of a given RBP are interconnected among themselves to what is expected theoretically and indicates the cohesiveness or local modularity of the network. Average value taken over all RBPs reflects the modularity of the network.	0.37	0.45
Betweenness	Betweenness centrality of a RBP measures the number of shortest paths between all pairs of RBPs in the network that pass through a RBP of interest – the higher the number of paths that pass through a RBP, the more important it is. Since this value depends on the total number of shortest paths in the network, it has to be normalized with the total number of RBP pairs in the network in order to compare networks (shown in braces).	55 (0.023)	45 (0.030)
Average path length	Average length of the shortest paths between all pairs of RBPs in the network.	2.72	2.65
Closeness	Closeness centrality is defined as the inverse of average length of all the shortest paths from a RBP of interest to all other RBPs in the network - note that closeness centrality defined this way implies that higher the closeness value, the higher the importance (centrality) of a RBP.	0.47	0.39
Diameter	The diameter of a network is the length of the longest path among all the shortest paths defined between two RBPs. It gives an estimation of the farthest distance between RBPs in the network.	6	5
Power law fit (exponent-alpha)	Fitting a power-law distribution function to the degree distribution of the network to study whether the network is likely to exhibit a scale-free network structure.	1.34	1.6

Table 2. Superposed FFL-FBL motifs identified in the post-transcriptional network of RBP-RBP interactions. In these 3-node motifs, node X regulates nodes Y and Z, node Y regulates Z and node Z in turn controls the expression of node X. Note that in these motifs there is a simultaneous feed-back (Z controls X) and feed-forward (X controls Z) activity in the same motif.

Node X	Node Y	Node Z
UBP3	SCP160	NOP56
MAP1	HEK2	NRD1
UBP3	BFR1	NOP56
SCD6	PUF2	PUF4
HEK2	PUF2	VTS1
BFR1	MAP1	HEK2
SCD6	BFR1	PUF4
SCD6	ARF3	PUF4
PAB1	GIS2	ARC15
SCP160	MAP1	HEK2
SCP160	PUB1	LYS1
SCP160	PUB1	BFR1
HEK2	PUB1	BFR1
HEK2	SCD6	BFR1
HEK2	UBP3	BFR1
HEK2	UBP3	SCP160
HEK2	PUF3	VTS1
MAP1	PUB1	PAB1
MAP1	PUB1	NRD1
NRD1	MRN1	NAB3
HEK2	LYS1	VTS1
HEK2	PUB1	VTS1

Table 3: Table showing the localization and degree of RBPs in the integrated network. Integrated network corresponds to the interactions between RBPs which interact both at the physical and post-transcriptional levels. Last two columns show the extent of overlap in the RNA targets identified by immunoprecipitation experiments for the two RBPs shown, as well as its significance calculated using a hyper-geometric model. Note that RBPs with comparable number of targets show a significant overlap in the number of targets and common localization.

Gene1	Alias	Localization	Degree in integrated network	Total targets	Gene2	Alias	Localization	Degree in integrated network	Total targets	Common targets	P-value
YJL080C	SCP160	ER	3	1337	YOR198C	BFR1	ER	5	1051	823	0
YNL251C	NRD1	Nucleus	3	997	YPL190C	NAB3	Nucleus	3	313	284	7.85E-219
YNL016W	PUB1	Cytoplasm	4	1639	YER165W	PAB1	Cytoplasm	9	1802	701	1.68E-68
YNL016W	PUB1	Cytoplasm	4	1639	YNL251C	NRD1	Nucleus	3	997	456	1.45E-64
YDR432W	NPL3	Nucleus	3	1153	YPL178W	CBC2	Nucleus	2	289	166	1.43E-59
YER165W	PAB1	Cytoplasm	9	1802	YPL178W	CBC2	Nucleus	2	289	176	2.99E-38
YER165W	PAB1	Cytoplasm	9	1802	YNL255C	GIS2	Cytoplasm	2	837	365	9.03E-34
YNL016W	PUB1	Cytoplasm	4	1639	YPL190C	NAB3	Nucleus	3	313	150	1.42E-22
YOR198C	BFR1	ER	5	1051	YGL014W	PUF4	Cytoplasm	5	534	127	7.58E-09
YIL160C	POT1	Ambiguous	1	35	YNL255C	GIS2	Cytoplasm	2	837	13	8.58E-05
YKL085W	MDH1	Mitochondrion	1	5	YGR192C	TDH3	Cytoplasm,Nucleus	1	161	3	0.000114
YNL016W	PUB1	Cytoplasm	4	1639	YDL167C	NRP1	Cytoplasm	1	7	6	0.000869
YGL014W	PUF4	Cytoplasm	5	534	YGR159C	NSR1	Nucleolus	1	560	59	0.002019
YJL080C	SCP160	ER	3	1337	YGL178W	MPT5	Cytoplasm	2	594	136	0.00213
YER165W	PAB1	Cytoplasm	9	1802	YNL004W	HRB1	Nucleus	2	40	18	0.003858
YDL078C	MDH3	Cytoplasm,Nucleus	1	56	YER165W	PAB1	Cytoplasm	9	1802	8	0.017165
YPR129W	SCD6	Cytoplasm	1	1709	YCL011C	GBP2	Nucleus	5	252	71	0.02105
YOR198C	BFR1	ER	5	1051	YPR042C	PUF2	Cytoplasm	1	658	95	0.042523
YLR244C	MAP1	Cytoplasm,Nucleus	2	201	YLR175W	CBF5	Nucleolus	2	156	8	0.044046
YOR198C	BFR1	ER	5	1051	YDR293C	SSD1	Cytoplasm	1	114	12	0.04515
YGL014W	PUF4	Cytoplasm	5	534	YHR209W	CRG1	NA	2	419	37	0.04534
YLR244C	MAP1	Cytoplasm,Nucleus	2	201	YCL011C	GBP2	Nucleus	5	252	11	0.050144
YCL011C	GBP2	Nucleus	5	252	YNL004W	HRB1	Nucleus	2	40	3	0.118456
YER165W	PAB1	Cytoplasm	9	1802	YKL130C	SHE2	Cytoplasm,Nucleus	1	24	4	0.121879
YCL011C	GBP2	Nucleus	5	252	YML117W	NAB6	Cytoplasm	1	54	0	0.137551
YGL178W	MPT5	Cytoplasm	2	594	YJR110W	YMR1	Cytoplasm	1	77	5	0.147462
YDR432W	NPL3	Nucleus	3	1153	YIR001C	SGN1	Cytoplasm	2	10	3	0.15171

YDR432W	NPL3	Nucleus	3	1153	YKL214C	YRA2	Cytoplasm,Nucleus	2	10	3	0.15171
YER165W	PAB1	Cytoplasm	9	1802	YIR001C	SGN1	Cytoplasm	2	10	4	0.154105
YCL011C	GBP2	Nucleus	5	252	YLR175W	CBF5	Nucleolus	2	156	6	0.163409
YER165W	PAB1	Cytoplasm	9	1802	YKL214C	YRA2	Cytoplasm,Nucleus	2	10	1	0.177334
YER165W	PAB1	Cytoplasm	9	1802	YGL245W	GUS1	Cytoplasm	1	12	2	0.223622
YGL014W	PUF4	Cytoplasm	5	534	YHR104W	GRE3	Cytoplasm,Nucleus	1	5	0	0.67293

Figure 1

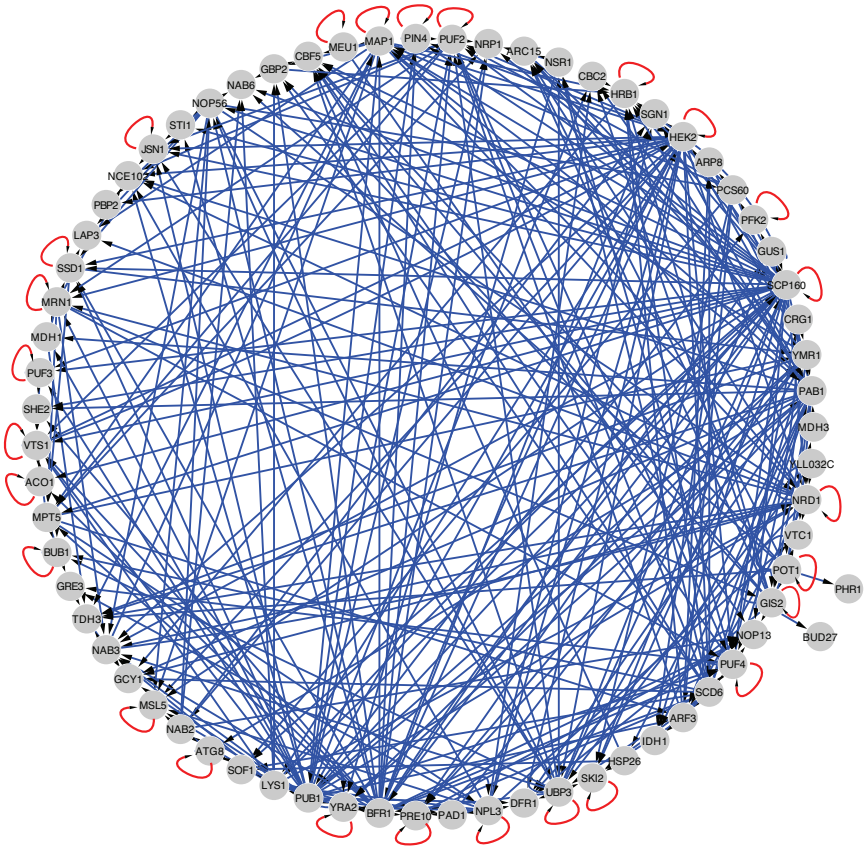
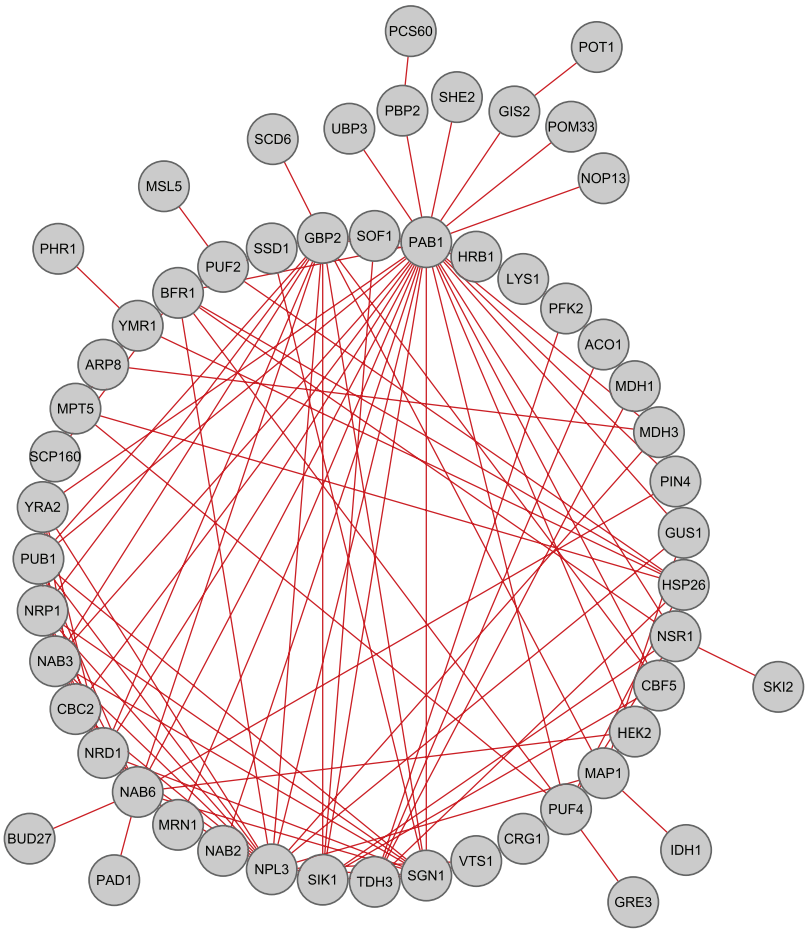

MEU1
MAP1
PIN4
PUF2
NRP1
ARC15
NSR1
CBC2
HRB1
SGN1
HEK2
ARP8
PCS60
PFK2
GUS1
SCP160
CRG1
YMR1
PAB1
MDH3
YLL032C
NRD1
VTC1
POT1
PHR1
GIS2
BUD27
NOP13
PUF4
FSCD6
ARF3
IDH1
HSP26
SKI2
UBP3
DFR1
NPL3
PAD1
PRE10
BFR1
YRA2
PUB1
LYS1
SOF1
ATG8
NAB2
MSL5
GCY1
NAB3
TDH3
GRE3
BUB1
MPT5
ACO1
VTS1
SHE2
PUF3
MDH1
MRN1
SSD1
LAP3
PBP2
NCE102
JSN1
STI1
NOP56
NAB6
GBP2
CBF5

Figure 2


PCS60
POT1
PBP2
SHE2
GIS2
POM33
UBP3
SCD6
NOP13
MSL5
PHR1
SSD1
GBP2
SOF1
PAB1
HRB1
LYS1
PUF2
BFR1
PFK2
YMR1
ACO1
ARP8
MDH1
MPT5
MDH3
SCP160
PIN4
YRA2
GUS1
PUB1
HSP26
NRP1
NSR1
NAB3
CBF5
SKI2
CBC2
HEK2
NRD1
MAP1
NAB6
PUF4
MRN1
CRG1
BUD27
IDH1
NAB2
VTS1
PAD1
NPL3
SIK1
TDH3
SGN1
GRE3