Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks

Nitish Mittala,b,1, Nilanjan Royb, M. Madan Babua, and Sarath Chandra Janga,a,1

aMRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom; and bDepartment of Biotechnology, National Institute of Pharmaceutical Education and Research, Sector 67, S. A. S. Nagar, Punjab 160062, India

Edited by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved October 5, 2009 (received for review June 22, 2009)

In eukaryotic organisms, gene expression requires an additional level of coordination that links transcriptional and posttranscriptional processes. Messenger RNAs have traditionally been viewed as passive molecules in the pathway from transcription to translation. However, it is now clear that RNA-binding proteins (RBPs) play an important role in cellular homeostasis by controlling gene expression at the posttranscriptional level. Here, we show that RBPs, as a class of proteins, show distinct gene expression dynamics compared to other protein coding genes in the eukaryote Saccharomyces cerevisiae. We find that RBPs generally exhibit high protein stability, translational efficiency, and protein abundance but their encoding transcripts tend to have a low half-life. We show that RBPs are also most often posttranslationally modified, indicating their potential for regulation at the protein level to control diverse cellular processes. Further analysis of the RBP-RNA interaction network showed that the number of distinct targets bound by an RBP (connectivity) is strongly correlated with its protein stability, translational efficiency, and abundance. We also note that RBPs show less noise in their expression in a population of cells, with highly connected RBPs showing significantly lower noise. Our results indicate that highly connected RBPs are likely to be tightly regulated at the protein level as significant changes in their expression may bring about large-scale changes in global expression levels by affecting their targets. These observations might explain the molecular basis behind the cause of a number of disorders associated with misexpression or mutation in RBPs. Future studies uncovering the posttranscriptional networks in higher eukaryotes can help our understanding of the link between different levels of regulation and their role in pathological conditions.

disease | posttranscriptional modifications | protein noise | regulation | systems biology

Gene expression is a highly regulated process and is controlled at several levels. In eukaryotes, control of gene expression first occurs at the level of transcription, where transcription factors regulate the synthesis of RNA of specific genes in response to different internal and external stimuli. On the other hand, at the protein level, several posttranscriptional modifications, such as phosphorylation by kinases and ubiquitin ligases, are known to spatially and temporally control the availability of functional protein products within the cell. However, a much less understood level of gene expression regulation, which occurs between these two layers, is due to the posttranscriptional control of RNAs. It is now increasingly known that this level is controlled by numerous factors with major players being the RNA-binding proteins (RBPs) (1–3) (see Fig. L4). Therefore, intricate coordination of regulation from these three different layers is important for finely controlling the flow of genetic information from genes to proteins in different conditions. Indeed, changes in gene expression due to aberrations at any of these three levels have been shown to be responsible for the cause of a number of disorders (4–8).

Development of DNA microarray technology has made it possible to measure the expression of each annotated gene at the transcript level. Indeed, this technique has been the high-throughput approach of choice to efficiently characterize the transcriptomes of several model organisms. One common assumption in DNA microarray experiments is that the level of mRNA of a particular gene reflects the amount of protein and there is little regulation at the posttranscriptional level. Recent studies comparing the high-throughput data for mRNA and protein abundances indicate that there is a very weak correlation between the number of transcripts and protein products of a gene, challenging this notion (9, 10). This suggests that the regulation of gene expression at the posttranscriptional level is predominant. For instance, in the eukaryotic pathogen, Trypanosoma cruzi, it is well known that gene expression is primarily controlled at the posttranscriptional level through RNA-binding proteins (RBPs) (11). These studies suggest the extensive role of posttranscriptional regulation in controlling gene expression in eukaryotes (12, 13).

In eukaryotes, transcription and translation occur in different compartments. This allows for a plethora of options to control RNA at the posttranscriptional level, including their splicing, polyadenylation, transport, mRNA stability, localization, and translational control (2, 3). Although some early studies revealed the involvement of RBPs in the transport of mRNA from nucleus to the site of their translation, increasing evidence now suggests that RBPs regulate almost all of the posttranscriptional steps shown in Fig. L4. For example, in humans, Nova protein is associated with splicing (14); PUF family proteins have been shown to play an important role during Caenorhabditis elegans oogenesis (15); Tap protein, like its yeast homolog Mex67, was reported as a bona fide mRNA nuclear export factor (16); Puf3p in yeast was shown to be responsible for localization of mitochondrial transcripts (17); and Pab1 was reported to regulate the initiation of translation (18). While the extensive role of RBPs in posttranscriptional control of cellular processes has been reviewed by several groups (1–3, 7), in yeast alone we found that RBPs are involved in multiple cellular processes and components (see Materials and Methods and SI Text S1).

All these aspects highlight the importance of RBPs in regulating gene expression at the posttranscriptional level.

Due to their central role in controlling gene expression at the posttranscriptional level, alteration in expression or mutations in either RBPs or their RNA targets (i.e., the transcripts which physically associate with the RBP) have been reported to be the cause of several human diseases such as muscular atrophies, neurological disorders, and cancer (6, 7, 19, 20). In particular, disorders such as myotonic dystrophy (DM) and oculopharyngeal muscular dystrophy (OPMD) have been attributed to RNA’s gain-of-function by CUG repeat expansion in the case of myotonic dystrophy protein kinase (DMPK) (19) and GCG repeat expansion in exon 1 of the RBP, PABPN1 in the case of OPMD (7), respectively. On the other hand, diseases caused by CUG repeat expansion in the case of myotonic dystrophy protein kinase (DMPK) (19) and GCG repeat expansion in exon 1 of the RBP, PABPN1 in the case of OPMD (7), respectively. On the other hand, diseases caused by CUG repeat expansion in the case of myotonic dystrophy protein kinase (DMPK) (19) and GCG repeat expansion in exon 1 of the RBP, PABPN1 in the case of OPMD (7), respectively.

Author contributions: S.C.J. designed research; N.M. performed research; N.M., N.R., M.M.B., and S.C.J. contributed new reagents/analytic tools; N.M. and S.C.J. analyzed data; and N.M. and S.C.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence may be addressed. E-mail: sarath@mrc-lmb.cam.ac.uk or nmittal@mrc-lmb.cam.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906940106/DCSupplemental.
hand, diseases like paraneoplastic opsoclonus-myoclonus ataxia (POMA) and spinal muscular atrophy (SMA) have been reported to be due to the RBPs loss of function (7), suggesting that mutations in either RBP or any of its interacting RNA target sequences can lead to extensive variations in their expression patterns and result in a number of diseases. In addition to the fitness defects that variations in RBPs can bring about in cells, it has been recently shown in yeast that RBPs form an important class of prionogenic proteins (21).

All these observations raise the questions: Are RBPs finely controlled in terms of their expression patterns and are there constraints on their expression patterns depending on the number of distinct RNA targets they control? To address this, in this study, we have analyzed the posttranscriptional network formed by RBPs in yeast, Saccharomyces cerevisiae at two distinct levels (Fig. 1B). The first involved asking whether RBPs as a group show distinct dynamic properties in comparison to non-RBPs in the whole genome. The second composed of understanding the constraints placed on dynamic properties of RBPs in relation to the number of distinct transcripts controlled by them. Our analysis at the first level revealed that RBPs, as a functional class, are rapidly turned over (i.e., less stable) at the transcript level and are tightly controlled at the protein level. Analysis of the posttranscriptional network formed by RBPs indicated that highly connected RBPs are more abundant and ubiquitously present within the cell.

Results

RBPs Show High Abundance and Tight Regulation at the Protein Level.

To compare and understand the differences in the gene expression dynamics of RBPs with other protein coding genes in S. cerevisiae, we first compiled the set of RBPs and non-RBPs as described in Materials and Methods (Fig. 1B). This allowed us to define a set of 561 proteins in yeast as those that encode for RNA-binding proteins and the remaining 5,685 proteins (from the complete set of protein coding genes) as non-RNA-binding proteins. We also collected high-throughput data documenting various dynamic properties of messenger RNA transcripts and their translated protein products in yeast from different sources as described in Materials and Methods. These properties included the mRNA stability, mRNA copy number, ribosome occupancy, protein stability, and abundance. In addition to these attributes of mRNAs and proteins, we also obtained the data describing the cell-to-cell variation in protein expression in a genetically homogenous population of cells, typically referred to as protein expression noise.

Messenger RNA half-life is a measure of transcript stability in the cell, whereas mRNA copy number reflects its abundance. We first asked whether RBPs as a functional class show a different tendency in comparison to non-RBPs in these properties. As a result of this analysis, we found that mRNAs encoding RBPs are significantly less stable (i.e., short half-life) at the transcript level compared to those genes that do not encode RBPs ($P = 3.1 \times 10^{-10}$, Wilcoxon test) (Fig. 2A). In yeast it has been shown that, in general, mRNAs of central physiological pathways have a longer half-life and mRNAs encoding regulatory and signaling proteins have a shorter half-life (22). In line with these observations, the observed lower half-life of RBPs in our analysis is consistent with their regulatory function and quick turn over at the transcript level. However, a comparison of the mRNA copy number of the two groups of genes, which is a proxy for mRNA abundance in the cell, indicated that RBPs are encoded by genes that exhibit much higher mRNA copy number ($P < 2.2 \times 10^{-10}$, Wilcoxon test) (Fig. 2B). Exclusion of translation and ribosome-associated genes, which form a significant fraction of the total repertoire of RBPs and are known to be highly expressed, did not change our results (SI Text S2). These observations suggest that RBPs tend to be less stable but more abundant at the transcript level, suggesting that abundance is a more prominent factor than their stability. Both mRNA half-life and mRNA abundance data indicate that RBP’s expression at mRNA level is likely to be transient but whenever they are transcribed they are produced at high concentrations.

Ribosome occupancy has been shown to be a measure of translational efficiency of mRNA. Higher ribosome occupancy relates to higher protein synthesis, and lower ribosome occupancy indicates low translation rate of mRNA. We next asked whether the ribosome occupancy i.e., rate of translation, of RBPs is higher than those for non-RBPs and whether their protein levels are higher.
within the cell. This analysis clearly revealed that RBPs have high ribosome occupancy \((P = 2.5 \times 10^{-12}, \text{Wilcoxon test})\) (Fig. 2C) and are also present in much higher concentrations \((P < 2.2 \times 10^{-16}, \text{Wilcoxon test})\) (Fig. 2D) with median abundances of RBPs being roughly double that observed for non-RBPs \((3,895 \text{ versus } 2,132)\). These results indicate that RBPs are abundant and are translated rapidly, supporting the versatile nature of their involvement in multiple posttranscriptional control mechanisms at different cellular locations \(\text{(SI Text S1)}\). Exclusion of ribosome and translation-associated factors from RBPs to compare nonribosomal RBPs against non-RBPs indicated that ribosomal RBPs contribute significantly to the observed differences in the rate of translation and protein abundance of RBPs \(\text{(SI Text S2)}\). Comparing the protein concentrations of nonribosomal RBPs with non-RBPs indicated that the former are still significantly more abundant \((P = 2.2 \times 10^{-2})\).

Stability of a protein measured as its half-life can be considered as a proxy for the life time of a protein in a cell. Therefore, to understand the degradation rates of RBPs and to compare them against non-RBPs we analyzed their protein half-lives \(\text{(see Materials and Methods)}\). This analysis revealed that RBPs are significantly more stable than non-RBPs, with RBPs exhibiting a median half-life of 71 min as against non-RBPs with 46 min \((P = 5 \times 10^{-12}, \text{Wilcoxon test})\) (Fig. 2E). Repeating the analyses with nonribosomal RBPs showed a consistent trend despite their exclusion \((P = 4.8 \times 10^{-2})\) \(\text{(SI Text S2)}\). Our observations on the increased protein stability and concentration of the RBPs compared to other proteins in the cell suggests that RBPs, whose main functional role is in the processing and localization of their mRNA targets, might be required at multiple subcellular locations and be used throughout the cell cycle. This may likely warrant their higher abundance and stability at the protein level. It is important to note that although RBPs exhibit high protein stability, they also show low transcript stability, which indicates that most RBPs that are stable at the protein level, might be avoiding cellular crowding of their transcripts by quick turnover at the transcript level. Indeed, it has been shown in yeast that most RBPs autoregulate their own activity at the transcript level \(\text{(23)}\).

To understand how these properties vary with different processes in which RBPs are involved, we divided RBPs into four major categories: translation, transport, RNA localization, and processing \(\text{using GO annotations and compared them with non-RBPs (SI Text S3)}\). This analysis revealed that the general trends observed for different categories are similar to those seen for RBPs as a whole although certain categories comprised relatively few RBPs.

Several RBPs have been shown to be posttranslationally modified, which adds a layer of flexibility to their function. Many of these posttranslational modifications have been shown to modify their RNA-binding properties or their subcellular localization. Indeed, at least four types of posttranslational modifications namely phosphorylation, ubiquitination, methylation, and SUMOylation have been reported for RBPs \(\text{(2)}\). High stability of RBPs indicates the potential that posttranslational modifications can offer in the diversification of their function. In fact, analysis of the number of kinase substrates in RBP and non-RBP populations using the currently available protein phosphorylation map for yeast \(\text{(24)}\), suggests that some kinases not only target higher number of RBPs compared to non-RBPs \((P = 2.7 \times 10^{-2})\) but also more kinases are associated with RBPs \((P < 2.2 \times 10^{-16})\) \(\text{(see SI Text S4)}\).

Gene expression is a highly dynamic process and because of its dynamic nature there is a large variation in a protein’s abundance among different cells in a population. This variation is termed as biological noise. Genes whose expression varies to a large extent show more noise and these are typically involved in stress response, amino acid biosynthesis, and heat shock. On the other hand, genes that show consistent expression during the cell cycle such as those involved in protein degradation and ribosomal proteins tend to show low noise \(\text{(25)}\). Here, we have explored this noise data, to address whether RBPs show significant difference from non-RBPs in terms of biological noise. As shown in Fig. 2F, RBPs were found to show significantly lower noise levels in comparison to non-RBPs \((P = 1.7 \times 10^{-12}, \text{Wilcoxon test})\). Reanalyzing the data by excluding ribosomal proteins still clearly indicated that RBPs exhibit much
lower noise compared to other protein coding genes ($P = 6.3 \times 10^{-6}$, Wilcoxon test) (SI Text S2). This analysis unambiguously reveals that low noise is an inherent property of all RBPs and suggests that RBPs are tightly regulated at the protein level with little variation in their expression from cell to cell. An independent analysis to compare the dynamic properties of RBPs with all of the protein coding genes (including RBPs) and varying the test statistic used to calculate the significance, did not change our results. This suggests that the trends observed are generally robust and are independent of the statistical test used (SI Text S5).

The Number of Distinct Targets Bound by a RBP Is Correlated with Its Cellular Abundance. RBPs are the key elements responsible for the posttranscriptional control of gene expression and when combined with their RNA targets, this information can be represented as a RBP-RNA network. Although, on a genomic scale, RBPs are believed to control diverse range of functions with some eukaryotic systems predominantly using posttranscriptional mechanisms for gene expression control (11, 13), large-scale elucidation of posttranscriptional networks is limited to few model organisms for a select set of RBPs. In yeast, few recent genomewide studies identified the targets for several RBPs using RIP-chip technology (23, 26). These studies revealed the important roles played by different families of RBPs and the structure of the posttranscriptional network formed by them. These high-throughput studies showed that the number of targets of a RBP can vary widely, from <10 to more than thousands. In this study we obtained this network, where nodes represent RBPs or their targets and links represent a distinct physical association between the RBP and the target RNA. We then systematically investigated the relationship between different dynamic properties of RBPs and the number of distinct RNA targets they control.

We first asked whether the number of targets of a RBP is correlated with its transcript stability by grouping the RBPs into different connectivity bins i.e., groups of RBPs comprising a number of distinct RNA targets (see Materials and Methods). As a result of this analysis, we found that there was a weak but positive correlation between them, suggesting that transcript turnover of RBPs may not be dependent on their number of targets ($R^2 = 0.18$, $P < 0.24$) (Fig. 3A). On the other hand, a comparison of the mRNA copy number of a RBP and its number of targets revealed a strong positive correlation between them, suggesting that RBPs with a high number of targets are likely to be more highly expressed at the mRNA level ($R^2 = 0.96$, $P < 1 \times 10^{-3}$) (Fig. 3B). For instance, PAB1 is a highly connected essential RBP that can bind to the poly(A) tail of an mRNA to regulate its translational initiation through its binding with eIF4G protein (18, 27). Indeed, it was reported to bind to 1,994 distinct RNA targets and was among the genes with very high mRNA copy number (7.1 mRNA copies/cell). These observations point to a direct link between the number of distinct targets of a RBP and its number of copies of mRNA in the cell. To test the existence of a correlation between the connectivity and the rate of translation or the absolute protein abundance profile of RBPs, we further explored the relationship between them (Fig. 3 C and D). This comparison uncovered a more general link between translational efficiency of a RBP and its degree. For instance, Pub1p is another poly(A) binding protein (28) that binds to diverse sets of transcripts involved in ribosome biogenesis, cellular metabolism, and transport (29). This protein was reported to be localized to both nucleus and cytoplasm (30). Hence to be present at different locations and to bind to a large number of transcripts it has to be translated more often and should be present in a higher number of copies. Consistent with this, we find that its transcript exhibits high ribosome occupancy. Indeed, Hogan et al. (23) demonstrated that RNA targets of highly connected RBPs were enriched for multiple processes and subcellular localizations. These results clearly unveil the strong relationship between the concentration of a RBP and the number of distinct RNA targets bound by them, indicating that RBPs responsible for controlling a wide range of targets must occur in a higher number of copies at the protein level. It is important to note that although RBPs as a group of genes are significantly higher expressed at the transcript and protein levels compared to non-RBP population, relative abundance of the RBPs is correlated to the hierarchy of a RBP, defined as the number of distinct RNA targets. It is also noteworthy to mention that the RBPs analyzed for connectivity in this section did not comprise core ribosomal proteins, strengthening the generality of these observations.

**RBPs Bound to Many RNA Targets Are Less Frequently Degraded and Tightly Controlled at Protein Level.** Although RBPs with a higher number of distinct targets are expressed at a higher level compared to those that control fewer targets, it is not evident whether their protein turnover rates hold a similar trend. Therefore, to understand whether there is any dependence between the stability of a RBP and the number of transcripts it controls, we used a similar
activity of RBPs with their noise value. As shown in Fig. 4 those that are poorly connected. Hence, we compared the connected RBPs can be expected to show less noise in comparison to non-RBPs and that previous studies reported that regulatory RBPs as a group show significantly lower noise in comparison to their misexpression or mutations in the sequences that are used for controlling gene expression of a wide range of genes. We also note that RBPs that are central to the cell are not only required in large quantities but are also found to be present for a longer time in the cell. All these observations suggest the importance of a posttranscriptional network of interactions in higher eukaryotes and raise several open questions in the regulation of gene expression beyond transcription. We believe that such questions could be addressed in the near future as more data from different levels of regulation become available.

Materials and Methods

Data on RNA-Binding Proteins in S. cerevisiae and Their Interactions. The complete list of annotated RBPs and the data for well-studied RBPs in S. cerevisiae were obtained from Hogan et al. (23). The total number of annotated RBPs in yeast reported in this study was 561 and mRNA targets for 41 RBPs have been systematically identified on a genome scale by employing the RIP-chip technology. This approach essentially consists of two steps. The first involves generation of two RNA samples, isolation of RBP-bound mRNA by immunoprecipitation of messenger-ribonucleoproteins using affinity purification, and isolation of cellular RNA representing the whole genome. The second step involves hybridization of the two isolated RNA samples using dual-color microarrays and are analyzed for enriched transcripts to detect the bound targets of a RBP (39). A total of 14,312 interactions comprising 41 RBPs and 5,025 genes in the entire genome of S. cerevisiae, which forms a network of posttranscriptional interactions between RBPs and the target RNAs obtained using this approach, were used in this study (23).

Data for Comparative Analysis of Expression Dynamics. To study the expression dynamics of RBPs in comparison to other protein coding genes in the genome and to analyze their relationship with the number of RNAs controlled by RBPs, we have used a variety of datasets. These include the transcript stability, mRNA copy number, ribosome occupancy, protein half-life, protein abundance, and protein noise. Transcript stability, which is measured as the mRNA half-life of a transcript, was obtained from Wang et al. (22) and contained mRNA half-lives for 4,687 genes in the entire genome. A key parameter describing the translational status of a gene is the fraction of its transcripts engaged in translation, which is defined by the ribosome occupancy (40). Likewise, the number of mRNA copies of a gene can be best described by the parameter mRNA copy number per cell. Both these parameters for genes in S. cerevisiae were obtained from Arava et al. (40) where the authors used velocity sedimentation to separate mRNAs bound to ribosomes
and quantified them using microarray analysis. mRNA copy number could be obtained for 5,643 genes whereas ribosome occupancy could be mapped for 5,700 genes, allowing us to study the extent of transcription abundance and translation rates of the genes and transcripts. Stability of a protein, which is an estimate of the duration it occurs within the cell is measured as the half-life of the protein. In yeast, protein half-lives have been estimated by Belle and coworkers for ~3,750 proteins by inhibiting translation (41). In this study, we used these data by excluding proteins whose half-lives have been obtained by extrapolation. Protein expression levels of the absolute number of protein molecules per cell, was obtained from Ghaemmaghami et al. (42). We could obtain abundance values for 3,868 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. (25). We could obtain noise data for 2,213 genes for cells grown on rich media. The authors in this study used two distinct measures for calculating protein noise, coefficient of variation (CV), which is the ratio of the standard deviation in the expression of a protein and its mean expression and distance from median (DM), which was calculated as the difference between the CV value of a protein and a running median of all CV values. In this study, we have used DM as a measure of protein noise as it was indicated to be a more robust measure compared to CV to understand protein-to-protein variations in noise levels (25). Since DM is the distance between the CV and median value of all CVs, negative values correspond to relatively less noise whereas positive values represent higher levels of noise in the protein expression.

Comparison of the Regulatory Properties of RBPs with Other Protein Coding Genes. To study whether RBPs show differences in dynamic properties when compared to other protein coding genes, we defined a non-RBP set of proteins. This set essentially comprised proteins in the whole genome after excluding the set of RBPs. To study whether RBPs show differences in dynamic properties when compared to other protein coding genes, we defined a non-RBP set of proteins. This resulted in five different bins corresponding to varying degrees of RBPs, done in such a way that each bin of RBPs contained roughly an equal number of proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory network of biological noise. Nature 441(7095):840–846.


37. Romansky S, De, B. Lang, Y. Kondo, and Y. Pilpel for providing helpful comments. This work was supported by MRC LMB (to N.M., S.C.J., and M.M.B.) and National Institute of Biomedical Research, Cambridge Commonwealth Trust (to S.C.J.).

38. Supplementary URL: http://www.mrc-lmb.cam.ac.uk/publications/senat/rbhp-dynamics

Acknowledgments. We thank members of the Theoretical and Computational Biology group at MRC Laboratory of Molecular Biology (LMB) for their feedback and helpful discussions during the early stages of this work. We also thank S. De, B. Lang, Y. Kondo, and Y. Pilpel for providing helpful comments. This work was supported by MRC LMB (to N.M., S.C.J., and M.M.B.) and National Institute of Biomedical Research, Cambridge Commonwealth Trust (to S.C.J.).

Supporting Information. For additional details relating to SI Text S1–S6, see Figures S1–S5, Tables S1–S3. An analysis of enrichment of cell-cycle regulated genes, performed to address the concerns of a reviewer, is detailed in SI Text S7.

Acknowledgments. We thank members of the Theoretical and Computational Biology group at MRC Laboratory of Molecular Biology for their feedback and helpful discussions. We also thank S. De, B. Lang, Y. Kondo, and Y. Pilpel for providing helpful comments. This work was supported by MRC LMB (to N.M., S.C.J., and M.M.B.) and National Institute of Biomedical Research, Cambridge Commonwealth Trust (to S.C.J.). Supplementary URL: http://www.mrc-lmb.cam.ac.uk/publications/senat/rbhp-dynamics


