

Post-transcriptional operons and regulons co-ordinating gene expression

Jack D. Keene & Patrick J. Lager

Center for RNA Biology, Department of Molecular Genetics & Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA; Tel: 001 919 684 5138; E-mail: keene001@mc.duke.edu

Key words: combinatorial control, decay regulons, gene expression, gene translation, proteomics, RNA-binding proteins, RNA splicing, RNP, transcriptomics

Abstract

Experiments reported over the past several years, including genome-wide microarray approaches, have demonstrated that many eukaryotic RNA-binding proteins (RBPs) associate with multiple messenger RNAs (mRNAs) both *in vitro* and *in vivo*. This multitargeted binding property of RBPs has led to a model of regulated gene expression in eukaryotes that we termed the post-transcriptional operon. This concept was established by an analogy between polycistronic mRNAs that are generated from bacterial operons, and the co-ordinated regulation of multiple monocistronic mRNAs by RBPs. Post-transcriptional operons represent a powerful mechanism to organize and express genetic information as functionally related combinations of monocistronic mRNAs. In fact, much of the diversification of individual proteomes may be determined by the combinatorial properties of post-transcriptional operons. This review examines data supporting the role of post-transcriptional operons and regulons in organizing genetic information and co-ordinating expression of functionally related transcripts from their origins at transcription to their subsequent splicing, export and translation.

The proteome is determined ultimately at the post-transcriptional level

The unique proteomes of each individual are the final determinants of its phenotype. The post-genomic era is addressing regulatory events that determine how proteomes and phenotypes are derived. While the gene expression profiles of individuals vary, their genomes are exceedingly similar to one another. The protein-coding genomes of individual humans are 99% identical to one another and those of other species such as mouse and chimpanzee are very similar also (Lander *et al.* 2001, Boguski 2002, Olsen & Varki 2004). The differences in phenotypes among individuals are most likely due to sequences involved in the regulation of expression of the genes

rather than in the sequences of the open reading frames themselves. Indeed, only about 1.2% of the human genome encodes proteins (Bejerano *et al.* 2004). Interestingly, the most ultraconserved regions of the human genome encode gene expression regulatory factors that are predominately RNA-binding proteins of the RNA recognition motif (RRM) family and transcription regulators such as POU proteins and the homeobox family (Bejerano *et al.* 2004). It is likely that post-transcriptional regulatory mechanisms including mRNA splicing, export, stability and translation, heavily influence the unique composition of an individual's proteome.

Recent studies in yeast and mammalian cells have reported a striking lack of correlation between the steady-state levels of mRNAs, as

determined using microarrays, and the proteins (i.e. proteomes) encoded by those mRNAs (Gygi *et al.* 1999, Ideker *et al.* 2001) These findings suggest strongly that post-transcriptional regulation at the level of mRNA translation is important for diversifying the proteome. Therefore, after the completion of transcription, post-transcriptional processes including splicing, export, RNA stability and translation, determine the expression profiles of proteins in eukaryotic cells. It is clear that RNA-binding proteins (RBPs) are responsible for much of this regulation but other mechanisms including microRNAs are similarly involved. We will describe our model of post-transcriptional regulation that is based on RBPs regulating multiple mRNAs to co-ordinate the production of macromolecular machines, such as the ribosome, the mitochondrion, and the neuronal synapse, as well as complex developmental events (Keene & Tenenbaum 2002, Hieronymus & Silver 2004).

Several studies have shown that RBPs associate with unique groups of mRNAs in the nucleus and cytoplasm each of which encodes a different protein involved in the same process (Levine *et al.* 1993, Gao *et al.* 1994, Buckanovich & Darnell 1997, Tenenbaum *et al.* 2000, Takizawa *et al.* 2000, Brown *et al.* 2001, Labourier *et al.* 2001, Lee & Schedl 2001, Eystathioy *et al.* 2002, Hieronymus & Silver 2003, Intine *et al.* 2003, Li *et al.* 2003, Liu *et al.* 2003, Ule *et al.* 2003, Waggoner & Liebhaber 2003, Chen *et al.* 2003, Inada & Guthrie 2004, Lopez de Silanes *et al.* 2004, Ryder *et al.* 2004, Gerber *et al.* 2004). For example, these include RBPs involved in pre-mRNA splicing (Ule *et al.* 2003), mRNA export (Hieronymus & Silver 2003) and in translational control (Jain *et al.* 1997, Antic *et al.* 1999, Mazan-Mamczarz *et al.* 2003; Table 1). Subsets of mRNAs that have been shown to be regulated by multi-targeted RBPs include those encoding early response functions, the translational apparatus (the ribosome), the mitochondrion, plasma membranes and synapses (Tenenbaum *et al.* 2000, Brown *et al.* 2001, Intine *et al.* 2003; Ule *et al.* 2003, Inada & Guthrie 2004, Gerber *et al.* 2004). In addition, mRNA subsets associated with specific RBPs have been identified that encode proteins and enzymes involved in regulatory functions such as chromatin modification and spindle body formation (Gerber *et al.* 2004). These findings support the previously posited theory of gene expression in which complex post-

transcriptional events are co-ordinately regulated by RBPs to produce functionally related proteins (Keene & Tenenbaum 2002; Figure 1). Furthermore, the model is consistent with the evolutionary expansion of multifunctionality of eukaryotic proteins and offers an explanation for how a limited number of mammalian genes can be sufficient to provide the proteomic complexity required to produce a multicellular organism. Therefore, the coordinated expression of mRNAs by RBPs at the post-transcriptional level can orchestrate complex functions much like transcription factors are believed to orchestrate gene expression at multiple promoters (Wen *et al.* 1998, Niehrs & Pollet 1999, Lockhart & Winzler 2000, Orphanides & Reinberg 2002, Bolouri & Davidson 2002). Taken together, these studies suggest the existence of a highly organized ribonucleoprotein infrastructure that may be as important in regulating the expression of eukaryotic genomes as the transcriptional apparatus itself (Keene 2001, Gerber *et al.* 2004). Future models of gene expression will need to accommodate overlapping transcriptional and post-transcriptional regulatory networks in order to explain the origins of complex traits in higher eukaryotes (Keene & Tenenbaum 2002, Ren *et al.* 2002, Hieronymus & Silver 2003, Intine *et al.* 2003, Ule *et al.* 2003).

Genes covalently linked in bacterial operons are not directly linked in eukaryotes

The search for direct linkages between human and mouse genes with related functions was undertaken following the discovery of operons in bacteria (Niedhardt & Savageau 1996, Judd 1998). Global analysis of gene expression in a number of eukaryotic systems has identified both co-ordinately and temporally linked (positive and negative time-delayed and inverted relationships) expression of functionally related genes termed synexpression groups (Niehrs & Pollet 1999, Qian *et al.* 2001). These concepts encompass both simultaneous and sequential interactions of regulatory components involved in gene expression. It was expected, and in some cases assumed, that functionally-related genes in higher cells that are co-ordinately regulated as in operons would be physically linked or at least localized to the same regions of mammalian chromosomes. However, neither operons nor polycistronic mRNA transcripts as they exist in

Table 1. List of recently demonstrated mRNA clusters identified in association with RNA-binding proteins in either yeast or mammalian cells that encode functionally related components of macromolecular complexes and cellular processes.

| Functional linkage | Regulation | RBP | Reference |
|--|----------------------------|------------------|---|
| Early response gene products | Stability & translation | ELAV/Hu | Gao 1994, Tenenbaum 2000, Lopez de Silanes 2004 |
| Ribosomal proteins and biogenesis | RP mRNA stability | La, Pub1p, Ccr4p | Intine 2003, Kenan & Keene 2004, Grigull 2004, Inada & Guthrie 2004 |
| Cell cycle components | Nuclear export | Yra1 | Hieronymus & Silver 2003 |
| Cell wall components | Nuclear export | Yra1/Mex67 | Hieronymus & Silver 2003 |
| Mitochondrial proteins (nuclear) | Translational repression | Puf3 | Gerber 2004 |
| Spindle pole body components | Translational repression | Puf5 | Gerber 2004 |
| Chromatin remodelling enzymes | Translational repression | Puf5 | Gerber 2004 |
| Cytoskeletal machinery | Translational repression | Puf5 | Gerber 2004, Hieronymus & Silver 2003 |
| Plasma membrane proteins | Translational repression | Puf1/2 | Gerber 2004 |
| Nucleolar regulatory components | Translational repression | Puf4 | Gerber 2004 |
| Inhibitory neuronal synapse | Splicing | Nova 1 | Ule <i>et al.</i> 2003 |
| Carbohydrate metabolic machinery | Nuclear export & stability | Yra1, Ccr4p | Hieronymus & Silver 2003 |
| Translation factors | Nuclear export | Mex67 | Hieronymus & Silver 2003, Grigull <i>et al.</i> 2004 |
| RNA-binding proteins | Nuclear export | Yra1/Mex67 | Hieronymus & Silver 2003 |
| Heat shock regulated proteins | Nuclear export | Yra1/Mex67 | Hieronymus & Silver 2003 |
| Germ line development | Translational repression | STAR/GLD1 | Lee & Schedl 2001, Ryder <i>et al.</i> 2004 |
| GW bodies | RNA turnover | GW182 | Eystathioy 2002 |
| Fragile X syndrome | Translation repression | FMRP | Brown <i>et al.</i> 2001, Chen <i>et al.</i> 2003, others |
| Cell polarity and fate determine | Localization | She2/She3 | Takizawa <i>et al.</i> 2000 |
| Membrane-associated factors | Membrane polysomes | Scp160p | Li <i>et al.</i> 2003 |
| Neuronal survival and apoptosis | Translation repression | Jerky | Liu <i>et al.</i> 2003 |
| Erythroid differentiation & myelogenous leukemia | mRNA stability | AlphaCP2 | Waggoner & Liebhaber 2003 |
| Iron metabolism | mRNA stability | Cth2, TTP | Puig <i>et al.</i> 2005 |

These RBPs have not been shown in all cases to co-ordinate production of the proteins encoded by the clustered mRNAs.

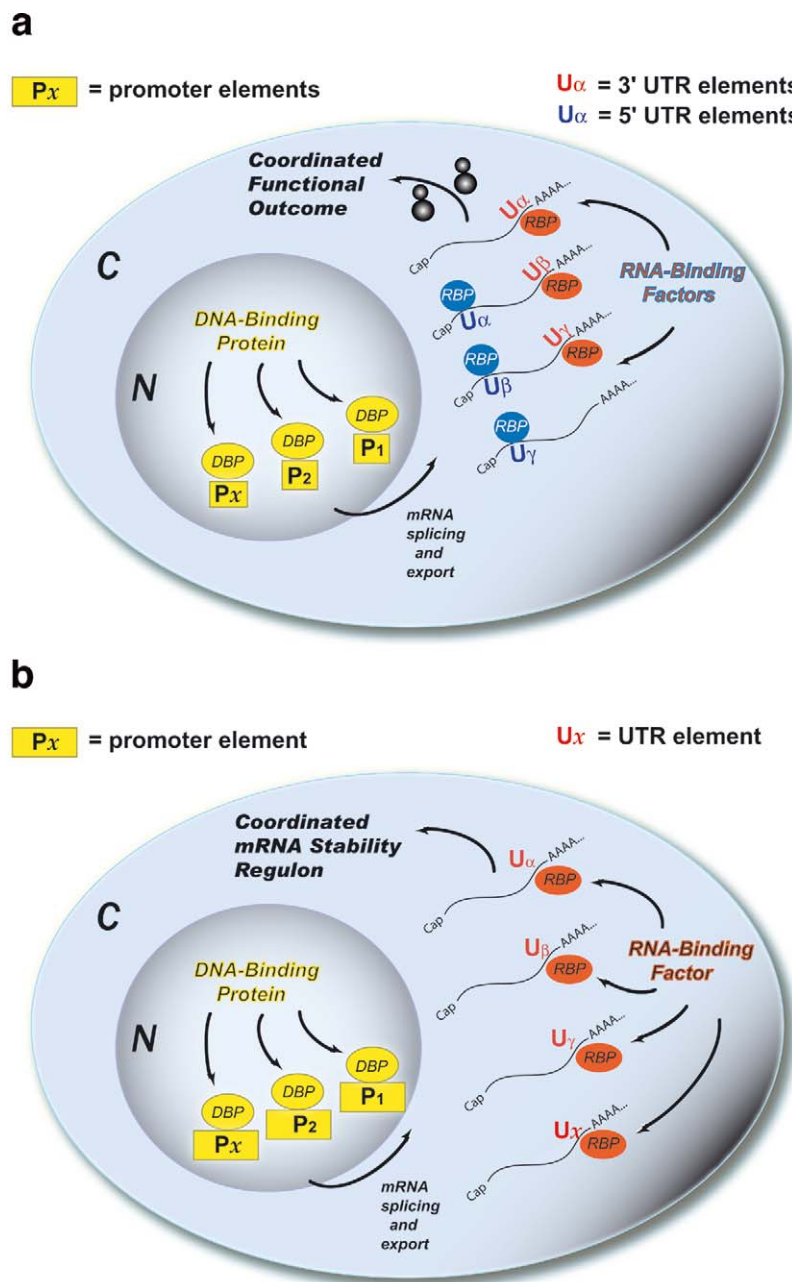


Figure 1. Depiction of dual gene expression networks localized to the nucleus (N) and cytoplasm (C). The nuclear networks (yellow) involve DNA-binding transcription factors, while the cytoplasmic networks (red and blue) function as post-transcriptional mRNA stability regulons (a) or post-transcriptional translational RNA operons (b). In the nucleus, multiple promoter elements on different genes can be regulated by transcription factors, while either 5' or 3'- untranslated regions (UTR elements in several different mRNAs) can be regulated in concert by post-transcriptional RNA-binding factors (RNA-binding proteins or micro RNAs). The post-transcriptional operon model proposes that multiple monocistronic mRNAs are co-ordinately regulated to generate functionally related proteins much like polycistronic bacterial mRNAs encode multiple tandem open reading frames. The co-ordinate regulatory principles depicted here apply also to pre-mRNA splicing, mRNA export and localization of multiple transcripts by *trans*-acting factors.

bacteria have been identified in higher eukaryotic cells. Gene expression 'neighborhoods' in *Drosophila* cells have been described in which an average of fifteen genes are coexpressed; however, they appear to have no functional relationship to one another (Spellman & Rubin 2002). In general, it is assumed by most investigators that gene expression is co-ordinated by transcription factors that work in combination to activate or repress multiple promoters of functionally-related genes (De La Brousse & McKnight 1993, Ren *et al.* 2002, Orphanides & Reinberg 2002, Bolouri & Davidson 2002). While this assumption may not prove to be as simple as hoped, the coordinated expression of mammalian genes based upon combinatorial recognition by transcription factors has been termed 'agglomerates' (Jacob 1997). The agglomerate model is based on the binding of multiple transcription factors to promoter elements of genes that encode functionally related proteins, presumably co-ordinating their expression in the nucleus (Figure 1). Thus, DNA-binding proteins act together combinatorially to activate or repress the transcription of each gene, and presumably multiple genes that orchestrate developmental processes and synexpression groups (Niehrs & Pollet 1999, Ren *et al.* 2002, Bolouri & Davidson 2002). The post-transcriptional operon model provides an alternative and yet complementary mechanism for co-ordinating gene expression of cellular genes that are otherwise dispersed across the chromosomes of eukaryotic cells (Keene & Tenenbaum 2002). Functional co-ordination by post-transcriptional operons or regulons can operate at splicing of premessenger RNA, mRNA export or mRNA localization. In Figure 1, mRNA stability regulons and translational regulons involving either 5' or 3' UTR elements are depicted. It is likely that transcriptional and post-transcriptional gene expression networks function together in concert and intercommunicate with one another on multiple levels to co-ordinate the production of the cell's proteome.

Post-transcriptional operons and regulons

The post-transcriptional operon model proposed that genes encoding functionally related proteins are regulated in a co-ordinate manner by the

association of specific subsets of messenger RNAs with RBPs (Keene & Tenenbaum 2002; Figure 2). The model also easily extends to post-transcriptional regulation mediated by small non-coding RNAs such as microRNAs (miRNAs). The term post-transcriptional operon refers to the physical linkage of functionally related mRNAs associated with specific RBPs or non-coding RNAs. This concept extends to eukaryotic regulons which are higher-order genetic units consisting of non-contiguous gene (mRNA) subsets under the control of a master regulatory gene (e.g. RBP). Decoupled transcription and translation in eukaryotic systems allows for multiple levels of post-transcriptional regulation at various steps of mRNA maturation including splicing, nuclear export, localization, stability and translation. This model is not only compatible with the more established paradigm of transcriptional co-ordination of gene expression, but also provides insights into the interaction of transcriptional and post-transcriptional gene expression networks necessary to carry out a given biological process. For example, ELAV/Hu proteins post-transcriptionally regulate mRNAs in the cytoplasm that produce early response gene products including regulatory transcription factors (Tenenbaum *et al.* 2000, 2002, Lopez de Silanes 2004).

The idea that RBPs may co-ordinate the expression of multiple mRNAs as distinct subsets was derived from experiments based upon *in-vitro* selection of RNAs that bind to the ELAV/HuB protein (Levine *et al.* 1993, Gao *et al.* 1994, Keene 1999, Brennan & Steitz 2001). ELAV/Hu RBPs are members of the RRM family that constitutes the largest known Pfam group of RBPs, and are among the most ultraconserved proteins known in the human genome (Lander *et al.* 2001, Bejerano *et al.* 2004). In fact, Hu RRM proteins are among the most ultraconserved proteins known to exist (Bejerano *et al.* 2004). Hu RBPs function in the stabilization and/or translational activation of target mRNAs (Jain *et al.* 1997, Fan & Steitz 1998, Antic *et al.* 1999, Keene 1999, Brennan & Steitz 2001, Mazan-Mamczarz *et al.* 2003). Work by Gao *et al.* (1994) demonstrated that *in-vitro* selection of RNAs from either randomized RNA libraries or from human brain 3' UTR mRNA libraries gave rise to multiple RNA species that contained AU-rich sequences. Interestingly, the

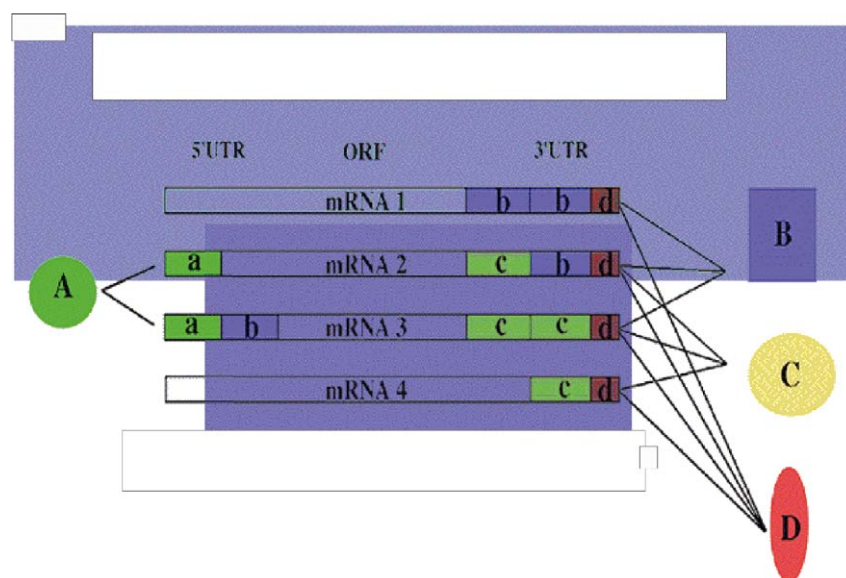


Figure 2. The Post-transcriptional Operon/Regulon model. Monocistronic mRNAs, each containing multiple 5' or 3' UTR elements (a, b, c, d) can interact with RNA-binding factors (A, B, C, D) including RNA-binding proteins or small regulatory RNAs that assist in co-ordinately regulating the expression of multiple mRNAs. Thus, mRNAs or pre-mRNAs that encode functionally related proteins, members of mRNA decay regulons, or cytoplasmically localized transcripts are defined by analogy to polycistronic mRNAs in bacteria which are collinear with the genes in operons on the DNA. These mechanisms provide combinatorial regulation of the genetic information *per se*, and the evolution of multifunctionality among eukaryotic proteins. For example, the occupation of a UTR element by one RBP or micro RNA may preclude adjacent UTR elements from binding other factors. Thereby, each mRNA may have several fates or functions depending on what proportion of the mRNA is occupied within a specific mRNP. Reprinted from *Mol Cell* 9: 1161, with permission of the publisher.

naturally derived 3'UTR mRNAs that were selected for binding to the ELAV/HuB protein *in vitro* represented a subset of brain transcripts, most of which were of the early response type that included transcription factors and other growth regulatory proteins including c-myc, c-fos, CREB2, and others (Levine *et al.* 1993, Gao *et al.* 1994). In fact, no single mRNA species dominated in these selected pools, suggesting that this RRM protein was able to bind to multiple species of messenger RNA with high specificity, yet the precise binding sequences were degenerate. Therefore, it was suggested that ELAV/Hu proteins could function as master regulatory factors that co-ordinate the expression of early response genes at the post-transcriptional level (Gao *et al.* 1994). In addition, it was hypothesized that ELAV/Hu proteins may play a role in the export of distinct subsets of early response gene transcripts from the nucleus, possibly with related functional outcomes (Keene 1999, 2001).

In-vivo analysis of mRNA populations associated with eukaryotic RBPs provided additional

findings critical to the origins of the post-transcriptional operon model: (1) RBPs associate with unique mRNA subpopulations, (2) a given mRNA species can associate with multiple RBPs, and (3) mRNA subpopulations associated with RBPs can be dynamic and combinatorial following biological perturbations (Tenenbaum *et al.* 2000, 2002, Keene & Tenenbaum 2002). These interactions are compatible with both simultaneous and sequential regulation of transcripts and the coupling of these processes from the nucleus to the cytoplasm characteristic of synexpression groups (Niehrs & Pollet 1999, Qian *et al.* 2001, Maniatis & Reed 2002, Keene & Tenenbaum 2002). Supporting data over the past several years have confirmed these findings in a number of eukaryotic systems including yeast (Hieronymus & Silver 2003, Gerber *et al.* 2004, Inada & Guthrie 2004), murine (Brown *et al.* 2001, Ule *et al.* 2003) and human (Intine *et al.* 2003, Arsham *et al.* 2003, Lopez de Silanes *et al.* 2004). Importantly, these studies have examined diverse families of RBPs that can

either positively or negatively affect target transcript expression, and expanded the known repertoire of mRNA subpopulations specifically regulated at the post-transcriptional level (Table 1). This has proved important to the identification of the numerous sequence elements within mRNA species necessary for *in-vivo* interaction with and regulation by RBPs.

Among the most striking data in support of post-transcriptional operons is that the half-lives of specific classes of mRNAs are co-ordinately regulated in mammalian cells and in *Saccharomyces cerevisiae* (Fan *et al.* 2002, Raghavan *et al.* 2002, Grigull *et al.* 2004, Wang *et al.* 2002, Yang *et al.* 2003, reviewed in Wilusz & Wilusz 2004, Rajasekhar & Holland, 2004 & Hieronymus & Silver 2004). Using genome-wide analysis of mRNA stability by genetically and chemically inactivating transcription in the presence of mutated mRNA stability factors, Tim Hughes' laboratory (Grigull *et al.* 2004) definitively demonstrated that mRNA decay operons were able to affect distinct classes of functionally related mRNAs *per* the post-transcriptional operon model (Keene & Tenenbaum, 2002). As noted by Grigull *et al.* (2004), their data were in agreement with previous data from Patrick Brown's and Daniel Herschlag's laboratories that also used genome-wide analysis of mRNA decay (Wang *et al.* 2002). In part, because the yeast system is so genetically tractable and accessible to genome-wide analysis and to gene ontology categorization, these RNA stability studies provided strong support for the post-transcriptional operon model using methods that were distinct from those used to detect mRNAs associated with regulatory RBPs (Tenenbaum *et al.* 2000, 2002). Therefore, a large number of studies using alternative methods have provided strong confirmation of the concept that the expression of functionally-related genes can be co-ordinated by multitargeted post-transcriptional processes (Hieronymus & Silver 2004).

The monocistronic mRNAs of eukaryotes contain conserved regulatory sequences found primarily in 3' and 5' untranslated regions (UTR) (reviewed in Keene & Tenenbaum 2002). The post-transcriptional operon model proposes that RBPs co-ordinately regulate mRNA subpopulations by interacting with transcripts containing shared sequence elements. Many mRNAs encoding both transcription factors and RBPs possess these regulatory elements suggesting

feedback mechanisms that interconnect transcriptional and post-transcriptional regulation. For example, the La RBP binds to mRNAs that together encode the components of the ribosome, a post-transcriptional machine, while the Hu RBPs bind to mRNAs that encode transcription factors. Thus, the 'crosstalk' between these two gene expression networks is likely to endow an interdependence that is synergistic. Given that many mRNAs contain multiple regulatory sequence elements, the model predicts that mRNA regulation is a result of the combinatorial effect of RBP interaction at both independent and mutually exclusive sequence sites (Figure 2). This allows for both spatial and temporal complexity in the expression of protein products, as well as the dynamics of associated mRNA subpopulations observed experimentally.

The combinatorial effect at the level of both RBP interaction and co-ordinated expression of dynamic mRNA subpopulations is a central concept of the eukaryotic post-transcriptional operon model. This provides for increased genetic complexity and agility at the post-transcriptional level while using a modest number of genes. It also provides for the evolution of protein multifunctionality by allowing transcripts to acquire new regulatory UTR sequence elements that provide an additional mode of expression that is independent of other mRNAs in the original clustered subset. Thus, an mRNA with a newly acquired UTR regulatory element could become a member of a different post-transcriptional operon or regulon. In this manner, post-transcriptional mechanisms can organize genomic information as monocistronic transcripts, provide increased combinatorial complexity and allow functional flexibility of expressed proteins.

Implications of post-transcriptional operons for developing systems

As noted above, several laboratories have demonstrated that mammalian and yeast RBPs can bind to multiple mRNAs encoding components of macromolecular structures and pathways, *per* the post-transcriptional operon model (Table 1). However, some of the most interesting post-transcriptional mRNA clusters may be those in which the mRNAs encode proteins *not otherwise*

known to function together. In other words, mRNA targets that are identified in association with many RBPs may not have an obvious functional relatedness. For example, mRNAs recovered by immunoprecipitation of ELAV/Hu proteins encode early response gene mRNAs that may function together in complex processes of growth, differentiation or neuronal plasticity (Keene 1999, Antic *et al.* 1999, Tenenbaum *et al.* 2000, Lopez de Silanes *et al.* 2004). They may represent mRNA decay regulons that co-ordinately stabilize the mRNAs in the group that activates differentiation factors. Such developmental processes are complex in that they require the orchestrated expression of hundreds or thousands of proteins, not otherwise known to work together as a synexpression group (Niehrs & Pollet 1999). However, as post-transcriptional operons, these could represent new macromolecular components or pathways that have not been previously recognized using traditional methods of investigation. In this sense, the discovery of individual post-transcriptional operons as mRNP clusters may reveal novel combinations of proteins or mRNA decay regulons that encode functionally related proteins.

In studies by Gerber *et al.* (2004) in which the yeast Puf3 protein, a translational repressor, was shown to associate with 154 mRNAs, over 87% of the transcripts encode mitochondrial functions. In the course of their study, mitochondrial localization of 27 additional proteins that were among the Puf3 mRNA target set were discovered in a separate proteomic analysis (Hug *et al.* 2003). Thus, the association of subsets of mRNAs with an RBP may provide a new paradigm of functional annotation that could be useful for the discovery of new cellular structures and pathways.

Interestingly, many of the mRNAs that are clustered by Puf3 in yeast encode mitochondrial ribosomal proteins. It is logical to speculate that when mitochondria evolved from an acquired free-living organism such as a bacterium and its (mitochondrial) ribosomal protein genes were transferred to the cell nucleus, whole operons were transferred and integrated into the host DNA as a block. Over time, the individual genes became dispersed over the genome of the host but the mRNAs encoded by each mitochondrial ribosomal gene in the host nucleus acquired UTR elements that bound to RBPs like Puf3 to help co-ordinate their expression. This

would have provided a post-transcriptional operon that fulfilled the regulatory role of the DNA operons that were gradually being dispersed among the chromosomes. Likewise, operative promoter elements in these genes may have evolved along with new transcription factors to help co-ordinate the production of these important mitochondrial components.

The fact that many RRM proteins such as ELAV/Hu and transcription factors such as the homeodomain proteins are ultraconserved in chordate genomes suggests that these immutable protein-coding regions may represent 'hardwired' components of the gene expression infrastructure. Such highly stable components of the RNP infrastructure could provide stability to gene expression networks that then accommodate the dynamics of mRNA trafficking from the genome to the proteome (Keene 2001). Such mechanisms are compatible with the concept of post-transcriptional operons that regulate developmental events and synexpression groups (Niehrs & Pollet 1999, Qian *et al.* 2001, Keene & Tenenbaum 2002), as well as the involvement of miRNAs in developmental regulatory pathways (reviewed in Ambros 2004).

Evolutionary implications of post-transcriptional operons

During the evolution of metazoans, several components of RNA processing and regulation have increased dramatically. For example, the numbers of RBPs and the length and complexity of 3' and 5' untranslated regions (UTRs) of mRNAs have increased significantly from prokaryotes to eukaryotes (Lander *et al.* 2001, Keene 2001, Keene & Tenenbaum 2002). In addition, proteins have become increasingly multifunctional in eukaryotes, suggesting that multiple pathways of regulation may have evolved to control the production of a protein for multiple functions in time and space. Ironically, the number of protein-encoding genes has not increased among eukaryotes in proportion to the expected size and complexity of the organism. It is generally believed that the evolutionary expansion of regulatory elements and processes has allowed a

relatively limited number of genes to be used in multiple combinations to diversify the proteome. This includes the advent of alternative splicing, regulatory small RNAs, repeated sequence elements such as ALU, LINES and SINES, and UTR elements. The post-transcriptional operon model provides mechanistic explanations for how multiple transcripts can be co-ordinately regulated during splicing, export, stability, localization and translation by widely dispersed sequence elements.

As noted above, there are at least a dozen conserved families of RBPs that have been identified, including the large RRM family that is ultra-conserved in many cases. The ELAV/Hu RRM proteins radiated in evolution from one homologue in *Drosophila* that is involved in splicing, to four homologues in mammals that are involved in mRNA export, stabilization and translation of early response mRNAs encoding protooncogenes and cytokines (Antic & Keene 1997, Keene 1999, Brennan & Steitz 2001). The functions of neuronal HuB, HuC and HuD proteins in regulating the stability and translational activation of early response gene mRNAs has been documented, as has the role of the ubiquitously expressed form, HuA (HuR) (reviewed in Keene 1999, 2001, Brennan & Steitz 2001). In addition, the neuronal Nova 1 and Nova 2 KH motif-type RBPs appear to be involved in alternative splicing of multiple mRNAs encoding synapses in the nervous system, as well as binding to sequences in other processed mature mRNAs encoding neuronal proteins (Ule *et al.* 2003).

While in bacteria the ribosomal genes are clustered in operons, they are widely dispersed among the chromosomes of eukaryotic organisms. Intine *et al.* (2003) recently reported, and a broader literature supports, that the La RBP binds to mRNAs encoding the ribosomal proteins in mammalian cells. Similar data using the yeast system demonstrated that the La counterpart also interacts in a pull-down microarray experiment with ribosomal protein mRNAs (Inada & Guthrie 2004). Therefore, it appears that, in addition to regulating the production of small RNAs involved in translation, La protein binds to ribosomal protein mRNAs and may help co-ordinate production of the translational apparatus itself (Kenan & Keene 2004). However, direct evidence for co-ordinated

regulation of translation of these mRNAs has not been published. These findings are relevant to data demonstrating that the Puf3 RBP in yeast binds to a subset of mRNAs that encode the nuclear contribution to the mitochondrion, most predominantly of which are the mitochondrial ribosomal proteins (Gerber *et al.* 2004). However, La protein does not appear to associate with mRNAs expressed from the mitochondrial DNA. Because cellular ribosomal genes are dispersed widely among the eukaryotic chromosomes, their diversification must have also evolved in parallel with the advent of RBP regulation at the level of their multiple mRNAs *per* the post-transcriptional operon model (Keene & Tenenbaum, 2002). It is assumed that eukaryotic mitochondrial ribosomal genes that are expressed from the host cell DNA must have been introduced into the cell genome more recently in evolution.

One could imagine that the acquisition of cells by cells to become organelles, such as the mitochondrion, would have allowed eukaryotes to cannibalize their gene expression networks as they formed independent modules. The interaction between two or more modular gene expression networks that are compartmentalized within organelles (including the nucleus) would be expected to provide enormous combinatorial powers of interconnectivity. One can imagine that such multilevelness of gene expression would evolve over time to eliminate useless genes in the acquired organelle and it would lose its ability to function independently. Thus, the synergy of coexisting gene expression networks would logically represent a significant adaptive advantage to higher organisms, especially if each had its own co-ordination mechanisms. In general, the concept of RNP modules and the coupling of transcription to translation involving RBPs should be considered in the context of the appearance of the nuclear membrane during the evolution of eukaryotes. Thus, the evolution of compartmentalized post-transcriptional gene regulatory networks probably involved modular acquisition of 5' and 3' UTRs as well. Indeed, the number of RBPs exploded in evolution (Keene 2001), as did the numbers of UTR elements and corresponding protein multifunctionality (Wool 1996, Naora & Naora 1999, Jeffery 1999) could take advantage of the agility provided by post-transcriptional operons.

The systems-level approaches recently applied to investigating the RNP infrastructure have provided insight into emergent post-transcriptional properties (reviewed in Keene & Tenenbaum 2002, Wilusz & Wilusz 2004, Rajasekhar & Holland 2004, Mesarovic *et al.* 2004, Hieronymus & Silver 2004). Broadly defining RNP components, interactions, dynamics and functional outcomes are important first steps in discerning the role of post-transcriptional regulation in co-ordinating gene expression. This global perspective is critical to understanding the effects of linked gene regulatory networks on cellular function, organismal development and the evolution of biological complexity.

References

- Ambros V (2004) The functions of animal microRNAs. *Nature* **431**: 350–355.
- Antic D, Keene JD (1997) Embryonic lethal abnormal visual (ELAV) RNA-binding proteins involved in growth, differentiation, and post-transcriptional gene expression. *Am J Hum Genet* **61**: 273–278.
- Antic D, Lu N, Keene JD (1999) ELAV tumor antigen, Hel-N1, increases translation of neurofilament M mRNA and induces formation of neurites in human teratocarcinoma cells. *Genes Dev* **13**: 449–461.
- Arsham AM, Howell JJ, Simon MC (2003) A novel hypoxia-inducible factor-independent hypoxic response regulating target of rapamycin and its targets. *J Biol Chem* **278**: 29655–29660.
- Bejerano G, Pheasant M, Makunin I *et al.* (2004) Ultra-conserved elements in the human genome. *Science* **304**: 1321–1325.
- Boguski MS (2002) The mouse that roared. *Nature* **420**: 515–516.
- Bolouri H, Davidson EH (2002) Modeling transcriptional regulatory networks. *BioEssays* **24**: 1118–1129.
- Brennan CM, Steitz JA (2001) HuR and mRNA stability. *Cell Mol Life Sci* **58**: 266–277.
- Brown V, Jin P, Ceman S *et al.* (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**: 477–487.
- Buckanovich RJ, Darnell RB (1997) The neuronal RBP Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol Cell Biol* **17**: 3194–3201.
- Chen L, Yun SW, Seto J, Liu W, Toth M (2003) The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience* **120**: 1005–1017.
- De La Brousse FC, McKnight SL (1993) Glimpses of allostery in the control of eukaryotic gene expression. *Trends Genet* **9**: 151–154.
- Eystathioy T, Chan EKL, Griffith K, Tenenbaum ST, Keene JD, Fritzier MJ (2002) A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. *Mol Biol Cell* **13**: 1338–1351.
- Fan XC, Steitz JA (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the *in vivo* stability of ARE-containing mRNAs. *EMBO J* **17**: 3448–3460.
- Fan J, Yang X, Wang W, Wood WH, Becker KG, Gorospe M (2002) Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc Natl Acad Sci (USA)* **99**: 10611–10616.
- Gao F, Carson C, Levine TD, Keene JD (1994) Selection of a subset of mRNAs from 3'UTR combinatorial libraries using neuronal RNA-binding protein, Hel-N1. *Proc Natl Acad Sci (USA)* **91**: 11207–11211.
- Gerber AP, Herschlag D, Brown PO (2004) Extensive association of functionally and cytologically related mRNAs with PUF family RNA-binding proteins in yeast. *PLoS Biol* **2**: E79.
- Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals post-transcriptional control of ribosome biogenesis factors. *Mol Cell Biol* **24**: 5534–5547.
- Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720–1730.
- Hieronymus H, Silver PA (2003) Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat Genet* **33**: 155–161.
- Hieronymus H, Silver PA (2004) A systems view of mRNP biology. *Genes Devel* **18**: 2845–2860.
- Hug W.-KI, Falvo JV, Gerke LC *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature* **425**: 686–691.
- Ideker T, Thorsson V, Ranish JA *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929–934.
- Inada M, Guthrie C (2004) Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. *Proc Natl Acad Sci (USA)* **101**: 434–439.
- Intine RV, Tenenbaum SA, Sakulich AL, Keene JD, Marais RJ (2003) Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. *Mol Cell* **12**: 1301–1307.
- Jacob F (1997) The Operon-twenty-five years later. *C.R. Acad Sci Paris* **320**: 199–206.
- Jain RG, Andrews LG, McGowan KM, Pekala P, Keene JD (1997) Ectopic expression of Hel-N1, an RNA-binding protein, increases glucose transporter (GLUT1) expression in 3T3-L1 adipocytes. *Mol Cell Biol* **17**: 954–962.
- Jeffery CJ (1999) Moonlighting proteins. *Trends Biochem Sci* **24**: 8–11.
- Judd B (1998) Genes and chromomeres: A puzzle in three dimensions. *Genetics* **150**: 1–9.
- Keene JD (1999) Why is Hu where? Shuttling of early-response-gene messenger RNA subsets. *Proc Natl Acad Sci (USA)* **96**: 5–7.
- Keene JD (2001) Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. *Proc Natl Acad Sci (USA)* **98**: 7018–7024.

- Kenan DJ, Keene JD (2004) La gets its wings. *Nat Struct Mol Biol* **11**: 303–305.
- Keene JD, Tenenbaum SA (2002) Eukaryotic mRNPs may represent post-transcriptional operons. *Mol Cell* **9**: 1161–1167.
- Labourier E, Blanchette M, Feiger JW, Adams MD, Rio DC (2001) The KH-type RNA-binding protein PSI is required for *Drosophila* viability, male fertility, and cellular mRNA processing. *Genes Devel* **16**: 72–84.
- Lander ES, Linton LM, Birren B *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Lee MH, Schedl T (2001) Identification of *in vivo* mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Devel* **15**: 2408–2420.
- Lerner RA, Seiser RM, Zheng T *et al.* (2003) Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes. *RNA* **9**: 1123–1137.
- Levine TD, Gao F, Andrews L, King PH, Keene JD (1993) Hel-N1: an autoimmune protein with binding specificity for uridylylated-rich 3' untranslated regions of growth factor messenger RNAs. *Mol Cell Biol* **13**: 3494–3504.
- Li AM, Watson A, Fridovich-Keil JL (2003) Scp160p associates with specific mRNAs in yeast. *Nucl Acids Res* **31**: 1830–1837.
- Liu W, Seto J, Sibille E, Toth M (2003) The RNA binding domain of jerky consists of tandemly arranged helix-turn-helix/homeodomain-like motifs and binds specific sets of mRNAs. *Mol Cell Biol* **23**: 4083–4093.
- Lockhart DJ, Winzler EA (2000) Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.
- Lopez de Silanes I, Zhan M, Lal A, Yang X, Gorospe M (2004) Identification of a target RNA motif for RNA-binding protein HuR. *Proc Natl Acad Sci (USA)* **101**: 2987–2992.
- Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* **416**: 499–506.
- Mazan-Mamczarz K, Galban S, Lopez de Silanes I *et al.* (2003) RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. *Proc Natl Acad Sci (USA)* **100**: 8354–8359.
- Mesarovic MD, Sreenath SN, Keene JD (2004) Search for organizing principles of understanding in systems biology. *Syst Biol* **1**: 19–27.
- Naora H (1999) Involvement of ribosomal proteins in regulating cell growth and apoptosis: Translational modulation or recruitment for extraribosomal activity? *Immunol Cell Biol* **77**: 197–205.
- Niedhardt FC, Savageau MA (1996) Regulation beyond the operon. In: Neidhart FC ed. *Escherichia coli and Salmonella Cellular and Molecular Biology*. Washington, DC: ASM Press, pp 1310–1324.
- Niehrs C, Pollet N (1999) Synexpression groups in eukaryotes. *Nature* **402**: 483–487.
- Olson MV, Varki A, (2004) Genomics. The chimpanzee genome—a bittersweet celebration. *Science* **305**: 191–192.
- Orphanides G, Reinberg D (2002). A unified theory of gene expression. *Cell* **108**: 439–451.
- Qian J, Dolled-Filhart M, Lin J, Yu H, Gerstein M (2001) Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identified new, biologically relevant interactions. *J Mol Biol* **314**: 1053–1066.
- Puig S, Askeland E, Thiele D (2005) Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation. *Cell* **120**: 99–110.
- Raghavan A, Ogilvie R L, Reilly C *et al.* (2002) Genomic-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucl Acids Res* **30**: 5529–5538.
- Rajasekhar VK, Holland EC (2004) Postgenomic global analysis of translational control induced by oncogenic signaling. *Oncogene* **23**: 3248–3264.
- Ren B, Cam H, Takahashi Y *et al.* (2002) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Devel* **16**: 245–256.
- Ryder SP, Frater LA, Abramovitz DL, Goodwin EB, Williamson JR (2004) RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nature Struct Mol Biol* **11**: 20–28.
- Spellman PT, Rubin GM (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol* **1**: 5–10.
- Takizawa PA, DeRisi JL, Wilhelm JE, Vale RD (2000) Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. *Science* **290**: 341–344.
- Tenenbaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci (USA)* **97**: 14085–14090.
- Tenenbaum SA, Lager PJ, Carson CC, Keene JD (2002) Ribonomics: Identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* **26**: 191–198.
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212–1215.
- Waggoner SA, Liebhaber SA (2003). Identification of mRNAs associated with alphaCP2-containing RNP complexes. *Mol Cell Biol* **23**: 7055–7067.
- Wang Y, Liu CXL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO (2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci* **99**: 5860–5865.
- Wen X, Fuhrman S, Michaels GS *et al.* (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci (USA)* **95**: 334–339.
- Wilusz CJ, Wilusz J (2004) Bringing the role of mRNA decay in the control of gene expression into focus. *Trends Genet* **20**: 491–497.
- Wool IG (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci* **21**: 164–165.
- Yang E, van Nimwegen E, Zavolan M *et al.* (2003) Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Res* **13**: 1863–1872.