Published in final edited form as:

Plant Sci. 2013 May ; 0: 55–62. doi:10.1016/j.plantsci.2013.01.009.

Review: Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants

Ian M. Silverman^{1,2,3,5}, Fan Li^{1,2,4,5}, and Brian D. Gregory^{1,2,3,4,*}

¹Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

²PENN Genome Frontiers Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

³Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

⁴Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

The eukaryotic transcriptome is regulated both transcriptionally and post-transcriptionally. Transcriptional control was the major focus of early research efforts, while more recently post-transcriptional mechanisms have gained recognition for their significant regulatory importance. At the heart of post-transcriptional regulatory pathways are *cis*- and *trans*-acting features and factors including RNA secondary structure as well as RNA-binding proteins and their recognition sites on target RNAs. Recent advances in genomic methodologies have significantly improved our understanding of both RNA secondary structure and RNA-binding proteins and their regulatory effects within the eukaryotic transcriptome. In this review, we focus specifically on the collection of these regulatory moieties in plant transcriptomes. We describe the approaches for studying RNA secondary structure and RNA-protein interaction sites, with an emphasis on recent methodological advances that produce transcriptome-wide datasets. We discuss how these methods that include genome-wide RNA secondary structure determination and RNA-protein interaction site mapping are significantly improving our understanding of the functions of these two elements in post-transcriptional regulation. Finally, we delineate the need for additional genome-wide studies of RNA secondary structure and RNA-protein interactions in plants.

© 2013 Elsevier Ireland Ltd. All rights reserved.

*Correspondence should be addressed to: Brian D. Gregory, Department of Biology, University of Pennsylvania, 433 S. University Ave., Philadelphia, PA 19104, Telephone: (215) 746-4398, Fax: (215) 898-8780, bdgregor@sas.upenn.edu.

⁵These authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

RNA secondary structure; RNA-binding proteins; post-transcriptional regulation; RNA genomics; RNA silencing; small RNAs

1. INTRODUCTION

Post-transcriptional regulation of the eukaryotic transcriptome can occur at any step of the RNA “life cycle” including maturation (e.g. splicing, polyadenylation, etc.); transport from the nucleus; localization within subcellular compartments; molecule stability; as well as the initiation, elongation, and termination of protein translation. Numerous *cis*- and *trans*-acting features and factors are integral to all of these regulatory processes.

The intrinsic secondary structure of RNA molecules is one such *cis*-acting feature. Secondary structure is the collection of intricate folding patterns that an RNA molecule forms through specific base pairing interactions encoded within its primary sequence [1–4]. Many RNA molecules cannot properly function without the formation of an extremely precise secondary structure [1–4]. For instance, ribosomal RNAs (rRNAs) must form structural folds that enable interactions with the correct ribosomal subunits at specific locations along their length, thereby allowing the formation of functional ribosomes [5]. Additionally, the structure of long non-coding RNAs (lncRNAs), not their primary sequence, drives their function in regulating gene expression [6, 7]. Structural elements also affect the overall steady state abundance and stability of many eukaryotic mRNAs [8–10]. Thus, the secondary structure of RNAs is often required for their functionality in diverse cellular and regulatory processes.

RNA-binding proteins are a group of *trans*-acting regulatory factors that are integral to the post-transcriptional regulation of eukaryotic transcriptomes. A cellular RNA is involved in a multitude of complex interactions with numerous *trans*-acting RNA-binding proteins from the initial processing of a transcript in the nucleus to its final translation and decay in the cytoplasm [11–13]. These RNA-binding proteins interact with mRNAs and form dynamic multi-component ribonucleoprotein complexes, which can be the functional forms of mRNAs [14]. It is only through their proper formation that transcripts are correctly regulated and precisely produce the required amount of protein in a eukaryotic cell [2, 11, 13, 14]. Thus, RNA-protein interactions are necessary for the functionality, processing, and regulation of many RNA molecules in plant cells. However, there is still much to be discovered about plant RNA-binding proteins and their interactions with target RNAs.

In this review, we explain the current understanding of the regulatory functions of RNA secondary structure and RNA binding proteins in plant transcriptomes. We discuss advances in genomic technologies that allow a more comprehensive view of plant RNA secondary structure and RNA-binding proteins and their functional significance. Finally, we delineate experiments that will improve the understanding of these post-transcriptional regulatory elements and their effects on plant transcriptomes.

2. RNA SECONDARY STRUCTURE

RNA secondary structure is critical to post-transcriptional gene regulation. For example, riboswitches are a potent class of gene regulatory structural elements within an mRNA that directly bind to small metabolites. Other post-transcriptional processes such as protein translation [15, 16] and RNA-mediated silencing [17, 18] are also tightly controlled by structural features within the RNA transcript. In the next few sections, we summarize the role of RNA structure in numerous regulatory processes.

2.1. Riboswitches

Riboswitches are regulatory elements located within an mRNA that bind directly to small molecule ligands with no requirement for a protein partner. They mediate gene expression via ligand-induced conformational changes [19, 20]. The thiamine pyrophosphate riboswitch was first discovered in bacteria [21, 22] and it can easily be distinguished by homology throughout the plant kingdom where it is located in the 3' untranslated region of the *thiaminC* (*THIC* (AT2G29630)) gene that is required for thiamine biosynthesis [23, 24]. The riboswitch controls the levels of mRNAs from this gene through conformational changes that either mask or unmask a 5' splice site, and is one of the most thoroughly studied examples of gene regulation by RNA secondary structure in plants.

There are likely many other riboswitches encoded by plant transcriptomes that do not have bacterial orthologs (our unpublished results). This is of note because the search for additional riboswitches in plants has been primarily based on sequence homology to known bacterial riboswitch domains [25], and is limited by the lack of currently available data on RNA secondary structure in the plant lineage. Several novel genomics approaches (see below) could prove useful in detection of additional plant riboswitches by revealing the secondary structure of thousands of mRNAs in a single experiment.

2.2. A role for secondary structure in translation

RNA secondary structure also plays an important role in translational regulation. This was first demonstrated by studies in which stable stem-loop structures were shown to strongly modulate protein yield of plasmid-encoded preproinsulin [15, 16]. A similar relationship between computationally predicted secondary structure and translational efficiency for both chloroplast and nuclear mRNAs was demonstrated in green algae and higher plants [26, 27]. DNA microarrays have been used to assess translational regulation genome-wide in *Arabidopsis* under a variety of stress conditions including dehydration [28], sucrose starvation [29], salinity, and temperature [30]. These studies revealed a significant inverse correlation between computationally predicted thermodynamic stability and ribosome loading in the 5' untranslated regions [31], suggesting that secondary structure could be important in mediating translational activity. Similarly, a strong correlation between genome-wide profiles of ribosome density and folding energy was identified in yeast [32], and a tendency for decreased secondary structure near the start codon of genes was observed for most cellular organisms and viruses [33]. In total, these results suggest that RNA secondary structure has a significant regulatory effect on overall protein translation from eukaryotic mRNAs.

2.3. RNA silencing pathways

Plant RNA silencing pathways are mediated by small RNAs, which consist primarily of microRNAs (miRNAs) and several classes of endogenous small interfering RNAs (siRNAs) [17, 34]. miRNAs are short ~21–22 nucleotide RNAs that direct post-transcriptional or translational repression of specific mRNAs through direct base pairing interactions with complementary sites in the target transcript sequence. The miRNA-target interaction is thought to extend along the entire length of plant miRNAs [35]. However, in animals, this interaction mostly involves complementary base pairing only between nucleotides 2 – 8 of a miRNA (counted from its 5' end) (seed region) and a binding site in a target transcript.

Various lines of evidence suggest that site accessibility mediates miRNA targeting efficiency [36–39]. The first study to incorporate target site structure in miRNA target prediction found 3-nucleotide accessible regions to be an important predictor of targeting efficiency in *Drosophila melanogaster* [38]. This observation was then extended to a more general trend of decreased structural complexity and increased accessibility in regions containing miRNA target sites [37]. A two-step nucleation/hybridization model based on Sfold-generated structure ensembles was used to identify contiguous 4-nucleotide accessible blocks as the best predictor for miRNA interaction efficiency [36]. A similar approach was used to explore the importance of target site and flanking region accessibility in miRNA targeting efficiency [39]. A recent genome-wide analysis of miRNA target site folding energies in four plant genomes revealed significantly higher site accessibility when compared with random sequences in genes rich in guanines and cytosines (GC-rich), but no such difference in GC-poor genes [40]. This observation is perhaps influenced by the expected contribution of high GC content to secondary structure [41], which in effect results in a lower average accessibility in GC-rich miRNA target genes. It is likely that this lower average accessibility necessitates selection of more unstructured sequences in the miRNA target site of these genes. Taken together, these findings suggest that miRNA site structure is a major determinant of their regulatory function in eukaryotes.

The majority of these studies has been performed in metazoan systems and consequently much less is known about the role of site accessibility in plant miRNA targeting efficiency. This is particularly important because both the targeting (e.g. full length versus seed region miRNA pairing) and regulatory (e.g. cleavage versus translational inhibition) mechanisms have been proposed to differ greatly between plants and animals [34, 42]. However, there is recent evidence that *Arabidopsis* miRNAs can direct both cleavage and translational inhibition regulatory mechanisms [34, 43, 44]. These findings lead to the intriguing possibility that target site accessibility could determine the functional fate of miRNA targets. In such a model, highly structured target sites would be unable to fully base pair with the effector miRNA, leading to translational inhibition, whereas unstructured or accessible sites would form the canonical full-length duplex and resolve via cleavage.

2.4. New insights into RNA secondary structure and its regulatory effects

Recently, several groups used high-throughput sequencing technologies to directly assay RNA secondary structure in the *Arabidopsis* transcriptome [10, 45], polyadenylated mRNAs from yeast [46], and non-coding RNAs from mouse [47] (Fig. 1). These transcriptome-wide

methods (parallel analysis of RNA structure (PARS), dsRNA-seq, ssRNA-seq, etc.) are all based on structure-specific chemical modification or nuclease digestions followed by high-throughput sequencing to provide more comprehensive maps of RNA secondary structure (Fig. 1). For instance, selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-seq) involves the addition of a chemical adjunct (NMIA or 1M7) to unpaired nucleotides of an RNA molecule, which blocks the extension of reverse transcriptase during subsequent primer extension reactions. The collection of primer extension reactions are sequenced and analyzed to determine the structures of the RNA(s) of interest [48]. dsRNA-seq uses a ribonuclease that selectively degrades single-stranded RNAs (RNaseONE) to select for base paired RNAs within the cellular population, which are then sequenced and analyzed. This approach was first used to characterize substrates of the RNA-dependent RNA polymerase RDR6 in *Arabidopsis* [45]. ssRNA-seq uses the same principle but a different nuclease (RNase V1) to capture the single-stranded cellular RNA population, and was recently used in conjunction with dsRNA-seq data to produce a comprehensive collection of secondary structure models in *Arabidopsis*, *Drosophila melanogaster*, and *Caenorhabditis elegans* based on a combination of experimental data and free energy minimization [9, 10]. Similarly, PARS involves the use of two ribonucleases (RNase S1 and V1) with different structural specificities to capture the double-stranded and single-stranded RNA populations, and was used to study the secondary structure of ~3000 yeast mRNAs [46]. Fragmentation sequencing (FragSeq) utilizes the single-stranded ribonuclease P1 and two background controls to comprehensively interrogate RNA secondary structure [47] (Fig. 1). In general, all four currently available methodologies are quite similar in nature, and are therefore readily applicable to studies of RNA secondary structure in various eukaryotic transcriptomes.

The results from these comprehensive approaches have uncovered numerous insights into secondary structure-mediated post-transcriptional regulatory processes. For instance, the average structural profiles of *Arabidopsis*, yeast, *C. elegans*, and *Drosophila* mRNAs revealed reduced base pairing in the regions around the start and stop codons [9, 10, 46]. These results support similar predictions based on free energy-based structure modeling [33, 49], and suggest that RNA secondary structure has a significant effect on translation from protein-coding transcripts in eukaryotic organisms. In support of this hypothesis, a recent study that integrated structure mapping data with ribo-seq, another sequencing-based approach for measuring ribosome-associated mRNAs using sucrose gradient centrifugation, revealed a strong positive correlation between increasing secondary structure and ribosome association [10]. This correlation could be a result of numerous mechanisms that are not mutually exclusive and include: 1) structure-induced stalling after translation initiation, such as during elongation and/or termination; and 2) enhanced ribosome association through structured regulatory elements. In total, transcriptome-wide structure mapping has revealed a significant regulatory link between RNA secondary structure and protein translation from eukaryotic mRNAs, including those of plants.

The data from these high-throughput approaches was also used to produce genome-wide profiles of miRNA target site structure for a number of eukaryotic species [9, 10]. From these analyses, a decreased propensity for base pairing was observed throughout the entire

length of 2,194 predicted miRNA binding sites in *Arabidopsis* as well as across the seed (miRNA nucleotides 2 – 8) paired regions in 16,221 *C. elegans* target sites. Furthermore, the average structure across these seed paired regions was found to be negatively correlated with *C. elegans* miRISC Argonaute protein (ALG-1) binding affinity [9, 33, 49]. Comprehensive identification of the AGO1-bound sites within miRNA target mRNAs is needed to examine this relationship in plants.

These new genomic approaches could have an exceptional impact on riboswitch discovery and functional characterization. Since riboswitches regulate gene expression via changes in their secondary (and tertiary) structure [19, 20], genome-wide maps of secondary structure could be used to identify putative riboswitch elements. Integration of these predictions with additional functional data such as protein and small molecule binding profiles would aid in distinguishing true riboswitches from metabolite-induced structural changes (such as a RNA-binding protein-directed conformational shift).

3. RNA-BINDING PROTEINS

RNA-binding proteins are a ubiquitous and heterogeneous class of proteins found in all organisms and characterized by the presence of one or more RNA-binding domains. RNA-binding proteins interact with single-stranded or double-stranded regions of RNA molecules through their binding domains, as well as with other cellular components through auxiliary domains. These proteins participate in nearly all aspects of post-transcriptional gene regulation, including both nuclear and cytoplasmic events.

RNA-binding domains are conserved from bacteria to humans, demonstrating their important roles in basic cellular processes. Two of the most widespread plant RNA-binding domains are the RNA Recognition Motif and the K Homology domain (see Table 1). The RNA Recognition Motif-containing protein family in particular is highly diversified within plants when compared to animals, with ~50% of these proteins having no obvious ortholog in metazoans [50]. Interestingly, this specificity of plant RNA-binding proteins and their roles in various stress responses suggest they might function primarily in plant exclusive processes [51]. Other RNA-binding domains include the cold-shock domain, dsRNA-binding domains, several types of zinc finger domains (the most abundant being C-x8-X-x5-X-x3-H), DEAD/DEAH box, pentatricopeptide repeat, Pumilio, and PIWI/Argonaute/Zwille (Table 1) [14, 51, 52]. RNA-binding proteins may contain a single binding domain (e.g. AtGR-RBP1), multiple copies of the same domain (e.g. AtPABP1), or a collection of different domains (e.g. AtSF1/BPP) [50]. These proteins also possess auxiliary domains that carry out a variety of functions, such as facilitating protein-protein interactions or acting as substrates for post-translational modifications. Glycine-rich and arginine-serine-rich domains are common auxiliary domains observed in plants and metazoans [53, 54].

The *Arabidopsis* genome encodes > 300 RNA-binding proteins (see TAIR release 10, Table 1, and [50, 51]), which is similar to most other plant species (see Table 1), as well as two metazoans (300 in *Drosophila melanogaster* [55] and 500 in *C. elegans* [56]). These estimates for plants are likely quite conservative due to a lack of experimentally validated RNA-binding proteins. For example, the PPR family of proteins in *Arabidopsis* has over 400

members, only 5 of which are currently annotated as RNA-binding proteins (TAIR release 10, Table 1). Therefore, the total number of RNA-binding proteins functioning in plant cells is potentially much higher, and may even approach the complexity seen in mammalian genomes with experimental estimates of well over 1,000 proteins (see Table 1, and [57, 58]). In the next few sections, we describe the recent advances and technologies that have greatly enhanced our understanding of the critical roles of RNA-binding proteins in post-transcriptional regulation, with a specific focus on the collections of these proteins found in the cytoplasm and nucleus. Furthermore, we discuss the importance of expanding these heretofore metazoan-specific approaches to plant systems.

3.1. RNA-binding proteins in post-transcriptional regulation

It is well established that RNA-binding proteins are intimately involved in nearly all aspects of eukaryotic post-transcriptional gene regulation. This is because mRNAs are accompanied by a dynamic suite of RNA-binding proteins that can determine their fate; from the initial processing and maturation of pre-mRNAs in the nucleus, to their eventual translation and decay in the cytoplasm. RNA-binding proteins involved in fundamental RNA processing and maturation processes are well conserved across all kingdoms [14, 51, 52]. For instance, the hnRNP and SR family of proteins are essential for both constitutive and alternative splicing in mammals [59, 60] and plants [61, 62]. Similarly, functional orthologs have been identified in both plants and animals for the cleavage and stimulation factors that are essential in polyadenylation. In the cytoplasm, RNA-binding proteins often determine the localization, translation, and stability of their target mRNAs. Although these latter functions for RNA-binding proteins have been more comprehensively studied in metazoans [52], a few well-described examples have also been reported in plants [63–65].

It seems that post-transcriptional regulation of transcript levels, stability, and translation is a significant mechanism by which plants rapidly reprogram their transcriptome and proteome in response to hormones and environmental stresses [51, 52, 64, 66, 67]. RNA-binding proteins are crucial to these post-transcriptionally regulated changes that are triggered in response to variable internal and external conditions and signals. For instance, abiotic stresses induce widespread changes in mRNA stability and translation in many plant species, changes that are likely regulated by RNA-binding proteins [68, 69]. This idea is supported by the observation that the expression and/or activity of these RNA-binding proteins is often regulated in response to many plant hormones (e.g. ABA) as well as environmental variables that include temperature, light, and salt stresses [51, 52, 64, 66, 67]. Many stress-activated RNA-binding proteins may function in plants as molecular chaperones that regulate the stability and translation of their bound RNAs in response to these different stress conditions [51, 52]. For instance, the ABA-induced phosphorylation of the RNA-binding protein AKIP1 in the fava bean (*Vicia faba*) is necessary for this protein to bind and regulate its target mRNAs [66]. Follow-up studies on the *Arabidopsis* orthologs of AKIP1, the UBA proteins, suggest that these proteins regulate both target mRNA stability and translation [63, 65].

RNA-binding proteins also have a key role in plant defense responses. The *Pseudomonas syringae* HopU1 effector protein modifies a number of *Arabidopsis* RRM-containing RNA-

binding proteins by ADP-ribosylation during infection by this bacterial pathogen. This modification reduced the ability of these RRM-containing proteins to bind and regulate their target RNAs, resulting in plants that are more susceptible to infection by this pathogen [70, 71]. These findings suggest that RNA-binding proteins are required for the proper response of plants to external biotic stressors through regulation of the defense transcriptome. Finally, the crucial role of RNA-binding proteins in adaptation to environmental stressors is also indicated by the aberrant growth phenotypes and increased or decreased hormone and/or stress responses in plants that over- or under-express specific RNA-binding proteins (e.g. AtNUP160) [64, 72–76]. These initial studies have provided a number of specific examples of RNA-binding proteins that regulate target RNA levels and translation during stress responses.

Detailed studies, mostly conducted in mammalian systems, have pointed to several recurring themes in RNA-binding protein-mediated regulation. First, RNA-binding proteins generally participate in multiple post-transcriptional processes and therefore, categorizing these proteins as splicing factors or stability factors should be done cautiously. A prominent example of an RNA-binding protein with multiple roles is SF2/ASF, which was originally identified as an essential splicing factor, and has now been implicated in translational control [77] and miRNA processing [78]. Second, a number of RNA-binding proteins, including the *Arabidopsis* PTB homologues as well as AtGRP7 and AtGRP8, bind and regulate their own mRNA [79, 80]. Finally, mRNAs interact with multiple RNA-binding proteins, and RNA-binding proteins in turn bind to functionally related sets of mRNAs, suggesting a combinatorial network for control of gene expression at the RNA level [81]. Thus, the final fate of an mRNA is determined by the entire complement of bound RNA-binding proteins. These themes in RNA-binding protein-mediated post-transcriptional regulation have yet to be fully substantiated in plant systems.

3.2. RNA-binding proteins and their RNA targets

A comprehensive analysis of bound RNA targets is necessary to understand the role of an RNA-binding protein in post-transcriptional gene regulation. This information is needed to determine the specific binding sites and sequence preferences (interaction motif(s)) of each RNA-binding protein. Initially, *in vitro* approaches were developed to identify RNA-binding protein interacting motifs. These include Electrophoretic Mobility Shift Assays (EMSAs), RNA-affinity chromatography, UV-crosslinking studies, and Systematic Evolution of Ligands by Exponential Enrichment (SELEX). Although these studies have proven useful in identifying RNA-binding protein interacting motifs and *cis*-elements, they are performed *in vitro* and thus may not reflect biologically relevant sequence specificities in cells.

For instance, EMSAs utilize *in vitro* binding and non-denaturing gel electrophoresis to identify changes in gel mobility due to binding events of protein and nucleic acids. While effective in demonstrating strong protein-nucleic acid interactions (especially DNA-protein), EMSAs may not be sensitive enough to capture weak or transient binding events. UV-crosslinking experiments, in which covalently linked RNA-protein complexes are interrogated by SDS polyacrylamide gel electrophoresis (SDS-PAGE), can be utilized to increase sensitivity with a concomitant loss in specificity. In RNA-affinity chromatography,

a specific RNA sequence is used to capture an interacting RNA-binding protein(s) from a total protein cell lysate. This approach is commonly used when trying to identify protein partners of known *cis*-regulatory sequences, but highly abundant or promiscuous RNA-binding proteins may confound results. Conversely, SELEX provides an approach to identify specific protein-interacting sequences for a particular protein of interest. SELEX reduces investigator bias but systematic biases may also exist due to the *in vitro* nature of the methodology. While all of these approaches can demonstrate an RNA-protein interaction, no one method represents the gold standard and the most reliable results are those confirmed by multiple methods.

More recently, *in vivo* approaches have been developed to more directly study RNA-protein interactions in eukaryotic cells (Fig. 2). For instance, RNA immunoprecipitation followed by RT-PCR, microarray (RIP-chip), or sequencing (RIP-seq) has been used extensively to identify mRNA targets of RNA-binding proteins from a variety of organisms [82]. RNA immunoprecipitation can also be performed in the presence of formaldehyde, which stabilizes interactions between RNAs and their interacting proteins, allowing for more stringent washing and elimination of RNA-binding protein association with non-biologically relevant targets after cell lysis. One caveat of this approach is that formaldehyde also crosslinks proteins to one another, and thus identified interactions may be indirect. However, revealing indirect associations may also be informative and biologically relevant given the complex nature of mRNPs in eukaryotic cells. The RNA immunoprecipitation approach has already been used to study the target mRNAs of a number of plant RNA-binding proteins [83–87].

A more specific approach for defining RNA-protein interactions is the Crosslinking and Immunoprecipitation (CLIP) approach. This approach relies on the crosslinking specificity of UV (254 nm) light, which covalently attaches RNAs to their interacting proteins (Fig. 2, [88]). An RNase digestion is performed during the isolation of the RNA-protein complexes, thereby revealing the specific interaction regions of RNA targets. This improves the resolution of CLIP to RNA-binding protein interacting sites in contrast to the whole RNA that is provided by RNA immunoprecipitation-based studies (Fig. 2). CLIP followed by high-throughput sequencing-based analysis of protein-bound RNA sites (HITS-CLIP) and several variant protocols (e.g. Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)) have been widely used to study RNA-binding proteins in a diverse set of metazoan cell types [89–91]. These methods reveal the entire complement of binding sites for a given RNA-binding protein, and have shed enormous insight into the role of RNA-binding proteins in pre-mRNA splicing [92, 93], stability [8], and translation [94]. The CLIP method has not been applied to study plant RNA-binding proteins to date, presumably due to the lack of cell-based models and specific immunoprecipitation-quality antibodies to biologically significant RNA-binding proteins. However, UV light has been shown to effectively crosslink mRNA and rRNA to interacting proteins in tobacco protoplasts and maize, respectively [95, 96], suggesting that this assay is applicable for studying plant RNA-binding protein-RNA interactions. Future CLIP-based studies will be necessary to uncover the interacting repertoires of numerous RNA-binding proteins and reveal the post-transcriptional regulatory significance of these proteins in plants.

Recently, several RNA-centric approaches for defining RNA-protein interaction sites have been developed. For instance, a photoactivatable-ribonucleoside enhanced crosslinking (PAR-CL) and oligo-dT affinity purification coupled with RNase (RNase I) digestion was used to comprehensively reveal the binding sites of RNA-binding proteins along mature mRNAs [57]. Our laboratory has also developed a similar method that utilizes differential nuclease digestion in the presence or absence of bound proteins without the need for oligo-dT affinity purification. We have used this approach to comprehensively map RNA-binding protein interaction sites transcriptome-wide in both plants and animals (termed protein interaction profile sequencing, PIP-seq) (see Fig. 2, and our unpublished results). These approaches are an RNA-centric means to define RNA-protein interaction sites across eukaryotic transcriptomes without the need for antibodies to specific proteins. Such approaches will likely be useful for studying the RNA-protein interactome of multiple plant species, cell types, etc.

4. CONCLUSIONS AND PERSPECTIVES

As highlighted herein, genomic era approaches have significantly increased our understanding of the post-transcriptional regulatory importance of RNA secondary structure and RNA-binding proteins in plants. However, there is still much to learn about these post-transcriptional regulatory elements and their effects on plant gene expression. For instance, most plant RNA-binding proteins have been identified on the basis of sequence homology to known RNA-binding domains, and the molecular function and targeting of many of these proteins *in vivo* is unclear. Therefore, we propose a two-fold strategy to increase our understanding of plant RNA-binding protein targets and functions. First, a comprehensive catalog of plant proteins with proven RNA-binding capability must be established. To do this, UV and/or formaldehyde-based crosslinking coupled to oligo-dT purification followed by mass spectrometry and high-throughput sequencing could be used to reveal mRNA-bound proteins [57] and protein-interacting sequences ([57], and our unpublished results) on a transcriptome-wide scale. Similar approaches have already been applied to study rice [97] and *Arabidopsis* [98] mRNA-interacting RNA-binding proteins on a limited scale. However, significantly more comprehensive, optimized, and standardized methods should be used across a broad spectrum of plant species. Second, the functional targets of specific RNA-binding proteins must be identified and characterized by immunopurification-based approaches (see Fig. 2). CLIP-based assays have been effective in mapping many metazoan RNA-binding protein-RNA interactions, and promise a wealth of information for plants. Application of these types of approaches to many plant species will undoubtedly provide a more complete understanding of the post-transcriptional regulatory mechanisms that affect the plant transcriptome.

Acknowledgments

This work was funded by NSF Career Award MCB-1053846 to B.D.G., NHGRI 5T32HG000046-13 to F.L., and NIGMS 5T32GM008216-26 to I.M.S.

Abbreviations

dsRNase	double-stranded RNase
ssRNase	single-stranded RNase
miRNA	microRNA
PIP-seq	protein interaction profile sequencing
siRNA	small interfering RNA
PARS	parallel analysis of RNA structure

References

- Buratti E, et al. RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol Cell Biol.* 2004; 24:1387–1400. [PubMed: 1472981]
- Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell.* 2009; 136:777–793. [PubMed: 19239895]
- Cruz JA, Westhof E. The dynamic landscapes of RNA architecture. *Cell.* 2009; 136:604–609. [PubMed: 19239882]
- Sharp PA. The centrality of RNA. *Cell.* 2009; 136:577–580. [PubMed: 19239877]
- Trappl K, Polacek N. The ribosome: a molecular machine powered by RNA. *Met Ions Life Sci.* 2011; 9:253–275. [PubMed: 22010275]
- Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* 2009; 106:9362–9367. [PubMed: 19474294]
- Khalil AM, Rinn JL. RNA-protein interactions in human health and disease. *Semin Cell Dev Biol.* 2011; 22:359–365. [PubMed: 21333748]
- Goodarzi H, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature.* 2012; 485:264–268. [PubMed: 22495308]
- Li F, et al. Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* 2012; 1:69–82. [PubMed: 22832108]
- Li F, et al. Regulatory Impact of RNA Secondary Structure across the Arabidopsis Transcriptome. *Plant Cell.* 2012; 24:4346–4359. [PubMed: 23150631]
- Keene JD. Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. *Proc Natl Acad Sci USA.* 2001; 98:7018–7024. [PubMed: 11416181]
- Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA.* 2012; 3:159–177. [PubMed: 22213601]
- Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet.* 2007; 8:533–543. [PubMed: 17572691]
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 2008; 582:1977–1986. [PubMed: 18342629]
- Kozak M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell.* 1980; 19:79–90. [PubMed: 7357609]
- Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci USA.* 1986; 83:2850–2854. [PubMed: 3458245]
- Baulcombe D. RNA silencing in plants. *Nature.* 2004; 431:356–363. [PubMed: 15372043]
- Jones-Rhoades MW, Bartel DP. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell.* 2004; 14:787–799. [PubMed: 15200956]
- Serganov A, Patel DJ. Molecular recognition and function of riboswitches. *Curr Opin Struct Biol.* 2012; 22:279–286. [PubMed: 22579413]

20. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol Cell*. 2011; 43:867–879. [PubMed: 21925376]
21. Mironov AS, et al. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*. 2002; 111:747–756. [PubMed: 12464185]
22. Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*. 2002; 419:952–956. [PubMed: 12410317]
23. Bocobza S, et al. Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev*. 2007; 21:2874–2879. [PubMed: 18006684]
24. Wachter A, et al. Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell*. 2007; 19:3437–3450. [PubMed: 17993623]
25. Hammann C, Westhof E. Searching genomes for ribozymes and riboswitches. *Genome Biol*. 2007; 8:210. [PubMed: 17472738]
26. Klaff P, Riesner D, Steger G. RNA structure and the regulation of gene expression. *Plant Mol Biol*. 1996; 32:89–106. [PubMed: 8980476]
27. Gallie DR. Posttranscriptional Regulation of Gene Expression in Plants. *Ann Rev Plant Physiol Plant Mol Biol*. 1993; 44:77–105.
28. Kawaguchi R, Girke T, Bray EA, Bailey-Serres J. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J*. 2004; 38:823–839. [PubMed: 15144383]
29. Nicolai M, et al. Large-scale analysis of mRNA translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control. *Plant Physiol*. 2006; 141:663–673. [PubMed: 16632591]
30. Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. Genome-wide analyses of early translational responses to elevated temperature and high salinity in *Arabidopsis thaliana*. *Plant Cell Physiol*. 2010; 51:448–462. [PubMed: 20089509]
31. Kawaguchi R, Bailey-Serres J. mRNA sequence features that contribute to translational regulation in *Arabidopsis*. *Nucleic Acids Res*. 2005; 33:955–965. [PubMed: 15716313]
32. Tuller T, Waldman YY, Kupiec M, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA*. 2010; 107:3645–3650. [PubMed: 20133581]
33. Gu W, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comp Biol*. 2010; 6:e1000664.
34. Voinnet O. Origin, biogenesis, and activity of plant microRNAs. *Cell*. 2009; 136:669–687. [PubMed: 19239888]
35. Rhoades MW, et al. Prediction of plant microRNA targets. *Cell*. 2002; 110:513–520. [PubMed: 12202040]
36. Long D, et al. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*. 2007; 14:287–294. [PubMed: 17401373]
37. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature*. 2005; 436:214–220. [PubMed: 15951802]
38. Robins H, Li Y, Padgett RW. Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci USA*. 2005; 102:4006–4009. [PubMed: 15738385]
39. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007; 39:1278–1284. [PubMed: 17893677]
40. Gu W, Wang X, Zhai C, Xie X, Zhou T. Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. *Mol Biol Evol*. 2012; 29:3037–3044. [PubMed: 22490819]
41. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 1981; 9:133–148. [PubMed: 6163133]
42. Axtell MJ, Westholm JO, Lai EC. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol*. 2011; 12:221. [PubMed: 21554756]
43. Brodersen P, et al. Widespread translational inhibition by plant miRNAs and siRNAs. *Science*. 2008; 320:1185–1190. [PubMed: 18483398]
44. Lanet E, et al. Biochemical evidence for translational repression by *Arabidopsis* microRNAs. *Plant Cell*. 2009; 21:1762–1768. [PubMed: 19531599]

45. Zheng Q, et al. Genome-Wide Double-Stranded RNA Sequencing Reveals the Functional Significance of Base-Paired RNAs in Arabidopsis. *PLoS Genet.* 2010; 6:e1001141. [PubMed: 20941385]
46. Kertesz M, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature.* 2010; 467:103–107. [PubMed: 20811459]
47. Underwood JG, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Meth.* 2010; 7:995–1001.
48. Lucks JB, et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci USA.* 2011; 108:11063–11068. [PubMed: 21642531]
49. Shabalina SA, Ogurtsov AY, Spiridonov NA. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 2006; 34:2428–2437. [PubMed: 16682450]
50. Lorkovic ZJ, Barta A. Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res.* 2002; 30:623–635. [PubMed: 11809873]
51. Lorkovic ZJ. Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 2009; 14:229–236. [PubMed: 19285908]
52. Ambrosone A, Costa A, Leone A, Grillo S. Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. *Plant Sci.* 2012; 182:12–18. [PubMed: 22118611]
53. Boucher L, Ouzounis CA, Enright AJ, Blencowe BJ. A genome-wide survey of RS domain proteins. *RNA.* 2001; 7:1693–1701. [PubMed: 11780626]
54. Mousavi A, Hotta Y. Glycine-rich proteins: a class of novel proteins. *Appl Biochem Biotech.* 2005; 120:169–174.
55. Lasko P. The drosophila melanogaster genome: translation factors and RNA binding proteins. *J Cell Biol.* 2000; 150:F51–56. [PubMed: 10908586]
56. Lee MH, Schedl T. RNA-binding proteins. *WormBook.* 2006:1–13.
57. Baltz AG, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell.* 2012; 46:674–690. [PubMed: 22681889]
58. Keene JD, Tenenbaum SA. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell.* 2002; 9:1161–1167. [PubMed: 12086614]
59. Choi YD, Grabowski PJ, Sharp PA, Dreyfuss G. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science.* 1986; 231:1534–1539. [PubMed: 3952495]
60. Zahler AM, Lane WS, Stolk JA, Roth MB. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* 1992; 6:837–847. [PubMed: 1577277]
61. Staiger D, Zecca L, Wiczeorek Kirk DA, Apel K, Eckstein L. The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.* 2003; 33:361–371. [PubMed: 12535349]
62. Lopato S, et al. atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes Dev.* 1999; 13:987–1001. [PubMed: 10215626]
63. Lambrmon MH, et al. UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants. *Mol Cell Biol.* 2002; 22:4346–4357. [PubMed: 12024044]
64. dit Frey NF, et al. The RNA binding protein Tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis*. *Plant Cell.* 2010; 22:1575–1591. [PubMed: 20484005]
65. Kim CY, Bove J, Assmann SM. Overexpression of wound-responsive RNA-binding proteins induces leaf senescence and hypersensitive-like cell death. *New Phytol.* 2008; 180:57–70. [PubMed: 18705666]
66. Li J, et al. Modulation of an RNA-binding protein by abscisic-acid-activated protein kinase. *Nature.* 2002; 418:793–797. [PubMed: 12181571]
67. Staiger D, Green R. RNA-based regulation in the plant circadian clock. *Trends Plant Sci.* 2011; 16:517–523. [PubMed: 21782493]

68. Branco-Price C, Kaiser KA, Jang CJ, Larive CK, Bailey-Serres J. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* 2008; 56:743–755. [PubMed: 18665916]
69. Park SH, et al. Post-transcriptional Control of Photosynthetic mRNA Decay under Stress Conditions Requires 3' and 5' Untranslated Regions and Correlates with Differential Polysome Association in Rice. *Plant Physiol.* 2012; 159:1111–1124. [PubMed: 22566494]
70. Fu ZQ, et al. A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature.* 2007; 447:284–288. [PubMed: 17450127]
71. Jeong BR, et al. Structure function analysis of an ADP-ribosyltransferase type III effector and its RNA-binding target in plant immunity. *J Biol Chem.* 2011; 286:43272–43281. [PubMed: 22013065]
72. Dong CH, et al. A putative *Arabidopsis* nucleoporin, AtNUP160, is critical for RNA export and required for plant tolerance to cold stress. *Mol Cell Biol.* 2006; 26:9533–9543. [PubMed: 17030626]
73. Kim JY, et al. Glycine-rich RNA-binding proteins are functionally conserved in *Arabidopsis thaliana* and *Oryza sativa* during cold adaptation process. *J Exp Bot.* 2010; 61:2317–2325. [PubMed: 20231330]
74. Lee BH, Kapoor A, Zhu J, Zhu JK. STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. *Plant Cell.* 2006; 18:1736–1749. [PubMed: 16751345]
75. Riera M, Redko Y, Leung J. *Arabidopsis* RNA-binding protein UBA2a relocates into nuclear speckles in response to abscisic acid. *FEBS Lett.* 2006; 580:4160–4165. [PubMed: 16828085]
76. Xiong L, et al. Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in *Arabidopsis*. *Dev Cell.* 2001; 1:771–781. [PubMed: 11740939]
77. Michlewski G, Sanford JR, Caceres JF. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell.* 2008; 30:179–189. [PubMed: 18439897]
78. Wu H, et al. A splicing-independent function of SF2/ASF in microRNA processing. *Mol Cell.* 2010; 38:67–77. [PubMed: 20385090]
79. Schoning JC, Streitner C, Meyer IM, Gao Y, Staiger D. Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Res.* 2008; 36:6977–6987. [PubMed: 18987006]
80. Stauffer E, Westermann A, Wagner G, Wachter A. Polypyrimidine tract-binding protein homologues from *Arabidopsis* underlie regulatory circuits based on alternative splicing and downstream control. *Plant J.* 2010; 64:243–255. [PubMed: 20735772]
81. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* 2008; 6:e255. [PubMed: 18959479]
82. Barkan A. Genome-wide analysis of RNA-protein interactions in plants. *Meth Mol Biol.* 2009; 553:13–37.
83. Streitner C, Koster T, Simpson CG, Shaw P, Danisman S, Brown JW, Staiger D. An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2012; 40:11240–11255. [PubMed: 23042250]
84. Crofts AJ, Crofts N, Whitelegge JP, Okita TW. Isolation and identification of cytoskeleton-associated prolamine mRNA binding proteins from developing rice seeds. *Planta.* 2010; 231:1261–1276. [PubMed: 20217123]
85. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods.* 2005; 37:376–386. [PubMed: 16314267]
86. Hafner M, Lianoglou S, Tuschl T, Betel D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods.* 2012; 58:94–105. [PubMed: 22926237]
87. Konig J, et al. iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *Journal of visualized experiments: JoVE.* 2011
88. Ule J, et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science.* 2003; 302:1212–1215. [PubMed: 14615540]

89. Huelga SC, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* 2012; 1:167–178. [PubMed: 22574288]
90. Sanford JR, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 2009; 19:381–394. [PubMed: 19116412]
91. Darnell JC, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell.* 2011; 146:247–261. [PubMed: 21784246]
92. Casati P, Walbot V. Crosslinking of ribosomal proteins to RNA in maize ribosomes by UV-B and its effects on translation. *Plant Physiol.* 2004; 136:3319–3332. [PubMed: 15466230]
93. Lorkovic ZJ, Wieczorek Kirk DA, Klahre U, Hemmings-Mieszczak M, Filipowicz W. RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. *RNA.* 2000; 6:1610–1624. [PubMed: 11105760]
94. Doroshenko KA, Crofts AJ, Morris RT, Wyrick JJ, Okita TW. RiceRBP: A Resource for Experimentally Identified RNA Binding Proteins in *Oryza sativa*. *Front Plant Sci.* 2012; 3:90. [PubMed: 22645600]
95. Schmidt F, et al. A proteomic analysis of oligo(dT)-bound mRNP containing oxidative stress-induced *Arabidopsis thaliana* RNA-binding proteins ATGRP7 and ATGRP8. *Mol Biol Rep.* 2010; 37:839–845. [PubMed: 19672695]
96. Delannoy E, Stanley WA, Bond CS, Small ID. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem Soc Trans.* 2007; 35:1643–1647. [PubMed: 18031283]
97. O'Toole N, et al. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol.* 2008; 25:1120–1128. [PubMed: 18343892]
98. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* 1992; 11:2655–2664. [PubMed: 1628625]
99. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science.* 1994; 265:615–621. [PubMed: 8036511]

Highlights

- RNA secondary structure is a potent *cis*-acting posttranscriptional regulatory moiety
- RBPs function in all posttranscriptional regulatory processes in eukaryotes
- Genomic studies are integral to the study of plant RBPs and RNA secondary structure

	SHAPE-Seq	PARS	FragSeq	ds/ssRNA-seq
RNA Input	<i>In vitro</i> nondenatured	<i>In vitro</i> renatured	<i>In vitro</i> renatured	<i>In vitro</i> renatured
Probe(s)	NMIA or 1M7	RNases S1 and V1	Nuclease P1	RNases ONE and V1
Control	DMSO-treated	None	Untreated and PNK-treated	None
Base pairing readout	Single-stranded only	Both	Single-stranded only	Both
Throughput	Limited by primer extension	Genome-wide	Genome-wide	Genome-wide
Examples	Synthetic RNA pool	polyadenylated mRNAs from yeast	non-coding RNAs from mouse	<i>Arabidopsis</i> , <i>Drosophila</i> , and <i>C. elegans</i> whole transcriptomes

Fig. 1.

The collection of genomic methodologies that are available for studying RNA secondary structure transcriptome-wide. All of these approaches are based on structure specific chemical modification (SHAPE-seq) or nuclease digestions (PARS, dsRNA-/ssRNA-seq, and FragSeq) followed by high-throughput sequencing to provide a comprehensive view of RNA secondary structure.

Method	Single RBP approaches			Global RBP approaches	
	RIP	CLIP	PAR-CLIP	PAR-CL	PIP-seq
Crosslinker	None or Formaldehyde	UV (254 nm)	UV (365 nm)	UV (365 nm)	UV (254 nm) or Formaldehyde
Specificity	Indirect	Direct	Direct	Direct	Direct or Indirect
Resolution	Whole transcript	Binding site	Single-nucleotide	Single-nucleotide	Binding site
Strategy	Stringent co-IP	Digestion and stringent co-IP	Digestion and stringent co-IP	Crosslinking and digestion	Crosslinking and differential digestion
Input types	Cells or Tissues	Cells or Tissues	Cells only	Cells only	Cells or Tissues
Use in Plants	Applicable	Not yet used, but likely applicable	Not yet used, but likely applicable	Not yet used, but likely applicable	Applicable
Diagram					
Examples	AtGRP7 [83], OsRBP-A [94], OsRBP-D [94]	Nova [88], hnRNPA1 [89], FMRP [91]	Pum2 [12, 86], QKI [12, 86], AGO2 [12, 86]	HEK-293T [57]	HeLa, foreskin fibroblasts, <i>P. patens</i> (our unpub. results)

Fig. 2.

The collection of genomic methodologies that are available for identifying the target mRNA repertoire of eukaryotic RNA-binding proteins. The single protein methods, RIP (RNA immunoprecipitation), CLIP (UV-crosslinking and immunoprecipitation), and PAR-CLIP (photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation), all require an antibody to a specific RNA-binding protein. The global RNA-binding approaches take advantage of ribonuclease insensitivity of protein-protected regions of RNAs to allow an unbiased, RNA-centric view of RNA-protein interaction sites.

Table 1

The total number of putative RNA-binding proteins containing each specified RNA-binding domain in four different eukaryotes.

Domain ¹	Arabidopsis ²	Rice ³	Maize ⁴	Human ⁵
RRM	197 (601)	22/180 (95/570)	285 (447)	597 (1012)
KH	28 (69)	3/26 (13/70)	53 (78)	113 (183)
CSD	5 (4)	2/3 (1/7)	4 (10)	18 (33)
DS-RBD	5 (30)	0/22 (0/42)	6 (19)	50 (114)
ZnF (C-x8-C-x5-C-x3-H)	5 (97)	0/40 (14/150)	60 (106)	64 (179)
DEAD/DEAH box	9 (150)	3/65 (81/211)	94 (70)	200 (409)
PPR⁶	450	1/477	303	8
RGG box⁷	56	17/170	86	152
PUF	25 (25)	0/15 (0/40)	22 (16)	8 (23)
PAZ	6 (20)	4/25 (37/48)	3 (6)	12 (27)
LSM	36 (75)	7/22 (9/52)	36 (55)	35 (64)

¹ Numbers are provided from various annotation databases as described below, as well as (in parentheses) from the InterPro database, using the following domains:

- RNA recognition motif domain (IPR000504)
- K Homology domain (IPR004087)
- Cold shock protein (IPR011129)
- Double-stranded RNA binding (IPR001159)
- Zinc finger, CCCH-type (IPR000571)
- DNA/RNA helicase, DEAD/DEAH box type, N-terminal (IPR011545)
- Pumilio RNA-binding repeat (IPR001313)
- Argonaute/Dicer protein, PAZ (IPR003100)
- Like-Sm (LSM) domain (IPR010920)

² Proteins from TAIR10 ('functional annotations' table) with the specified domain and RNA-binding function.

³ Proteins from RGAP7 ('locus info' and 'Pfam' tables) with the specified domain. The numbers of proteins that are found in the RiceRBP database by blastp search with an e-value cutoff of 1e-50 are also given (e.g. 22/180 means 22 of the 180 RRM domain-containing proteins are found in RiceRBP).

⁴ Proteins from Phytozome v8.0 ('annotation info' table) with the specified domain.

⁵ Proteins from Pfam (*Homo sapiens* proteome file) with the specified domain.

⁶ From [96, 97].

⁷ Determined by searching respective proteome sequences for at least 3 RGG occurrences within a span of 30 amino acids [98, 99].