11-2012

Regulatory Impact of RNA Secondary Structure across the *Arabidopsis* Transcriptome

Fan Li
University of Pennsylvania, fanli.gcb@gmail.com

Lee E. Vandivier
University of Pennsylvania, evlee@sas.upenn.edu

Matthew R. Willmann
University of Pennsylvania

Ying Chen
University of Pennsylvania

Brian D. Gregory
University of Pennsylvania, bdgregor@sas.upenn.edu

Regulatory Impact of RNA Secondary Structure across the *Arabidopsis* Transcriptome

Abstract

The secondary structure of an RNA molecule plays an integral role in its maturation, regulation, and function. However, the global influence of this feature on plant gene expression is still largely unclear. Here, we use a high-throughput, sequencing-based, structure-mapping approach in conjunction with transcriptome-wide sequencing of rRNA-depleted (RNA sequencing), small RNA, and ribosome-bound RNA populations to investigate the impact of RNA secondary structure on gene expression regulation in *Arabidopsis thaliana*. From this analysis, we find that highly unpaired and paired RNAs are strongly correlated with euchromatic and heterochromatic epigenetic histone modifications, respectively, providing evidence that secondary structure is necessary for these RNA-mediated posttranscriptional regulatory pathways. Additionally, we uncover key structural patterns across protein-coding transcripts that indicate RNA folding demarcates regions of protein translation and likely affects microRNA-mediated regulation of mRNAs in this model plant. We further reveal that RNA folding is significantly anticorrelated with overall transcript abundance, which is often due to the increased propensity of highly structured mRNAs to be degraded and/or processed into small RNAs. Finally, we find that secondary structure affects mRNA translation, suggesting that this feature regulates plant gene expression at multiple levels. These findings provide a global assessment of RNA folding and its significant regulatory effects in a plant transcriptome.

Disciplines

Amino Acids, Peptides, and Proteins | Biology | Nucleic Acids, Nucleotides, and Nucleosides

The Plant Cell, Vol. 24: 4346–4359, November 2012, www.plantcell.org

LARGE-SCALE BIOLOGY ARTICLE

Regulatory Impact of RNA Secondary Structure across the *Arabidopsis* Transcriptome[W][OA]

Fan Li,[a,b,c,1] Qi Zheng,[a,b,1] Lee E. Vandivier,[a,b,d] Matthew R. Willmann,[a,b] Ying Chen,[a,b,c] and Brian D. Gregory[a,b,c,d,2]

[a] Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104
[b] PENN Genome Frontiers Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104
[c] Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, Pennsylvania 19104
[d] Cell and Molecular Biology Graduate Program, University of Pennsylvania, Philadelphia, Pennsylvania 19104

The secondary structure of an RNA molecule plays an integral role in its maturation, regulation, and function. However, the global influence of this feature on plant gene expression is still largely unclear. Here, we use a high-throughput, sequencing-based, structure-mapping approach in conjunction with transcriptome-wide sequencing of rRNA-depleted (RNA sequencing), small RNA, and ribosome-bound RNA populations to investigate the impact of RNA secondary structure on gene expression regulation in *Arabidopsis thaliana*. From this analysis, we find that highly unpaired and paired RNAs are strongly correlated with euchromatic and heterochromatic epigenetic histone modifications, respectively, providing evidence that secondary structure is necessary for these RNA-mediated posttranscriptional regulatory pathways. Additionally, we uncover key structural patterns across protein-coding transcripts that indicate RNA folding demarcates regions of protein translation and likely affects microRNA-mediated regulation of mRNAs in this model plant. We further reveal that RNA folding is significantly anticorrelated with overall transcript abundance, which is often due to the increased propensity of highly structured mRNAs to be degraded and/or processed into small RNAs. Finally, we find that secondary structure affects mRNA translation, suggesting that this feature regulates plant gene expression at multiple levels. These findings provide a global assessment of RNA folding and its significant regulatory effects in a plant transcriptome.

INTRODUCTION

RNAs fold into intricate three-dimensional structures (Brierley et al., 2007; Montange and Batey, 2008), which are determined by specific base pairing interactions encoded within their primary sequences (Buratti et al., 2004; Cooper et al., 2009; Cruz and Westhof, 2009; Sharp, 2009). It is becoming increasingly clear that RNA function depends upon proper folding (Buratti et al., 2004; Cooper et al., 2009; Cruz and Westhof, 2009; Sharp, 2009). In fact, it is known that rRNAs must form intricate secondary structures to enable proper formation of functional ribosomes (Trappl and Polacek, 2011). More recently, it has been shown that the structure of long noncoding RNAs is most likely necessary for their function in gene expression regulation (Khalil and Rinn, 2011; Ulitsky et al., 2011; Wan et al., 2011). Thus, the secondary structure of RNAs is required for the functionality of these molecules.

Interestingly, there is also increasing evidence suggesting that eukaryotic pre-mRNA molecules need to fold into precise secondary structures during mRNA maturation processes (e.g., splicing and polyadenylation) (Buratti et al., 2004; Cooper et al., 2009; Cruz and Westhof, 2009; Sharp, 2009). For example, structural elements within the pre-mRNA transcript can either repress or aid splicing by masking or organizing splice sites, respectively (Raker et al., 2009; Warf and Berglund, 2010), and can also regulate polyadenylation (Klasens et al., 1998; Zarudnaya et al., 2003). Recently, it was also discovered that specific structural elements affect the overall stability of many eukaryotic mRNAs (Goodarzi et al., 2012). Thus, the secondary structure of mRNAs is necessary for proper maturation and regulation. However, the overall effects of RNA folding on these processes and gene expression regulation are largely unclear.

Intramolecular and intermolecular base pairing is also required for the evolutionarily conserved RNA silencing pathways of eukaryotes (Bartel, 2004; Baulcombe, 2004; Carthew and Sontheimer, 2009). RNA silencing is initiated by the production of small RNAs (smRNAs), which in *Arabidopsis thaliana* consist of microRNAs (miRNAs) and several classes of endogenous small interfering RNAs (siRNAs) (Baulcombe, 2004; Voinnet, 2009). miRNAs and siRNAs are generated from self-complementary fold-back structures or heteroduplex double-stranded RNA (dsRNA), respectively, through the activity of DICER or DICER-LIKE (DCL) RNase III-type ribonucleases (Bartel, 2004; Meister and Tuschl, 2004; Jones-Rhoades et al., 2006; Carthew and Sontheimer, 2009). Once bound by an ARGONAUTE protein in an RNA-induced silencing complex (RISC), plant smRNAs direct RISC

[1] These authors contributed equally to this work.
[2] Address correspondence to bdgregor@sas.upenn.edu.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Brian D. Gregory (bdgregor@sas.upenn.edu).
[W] Online version contains Web-only data.
[OA] Open Access articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.112.104232

complexes to silence targets posttranscriptionally through intermolecular base pairing by degrading and inhibiting the translation of target RNAs, as well as transcriptionally by the deposition of DNA methylation and heterochromatic histone modifications to specific loci (Almeida and Allshire, 2005; Tomari and Zamore, 2005; Voinnet, 2009). In plants, smRNA–target interactions involve extensive complementary base pairing along the entire length of the smRNA (Voinnet, 2009). Overall, base paired RNAs are at the core of both the biogenesis and function of *Arabidopsis* smRNAs, providing further evidence of the importance of RNA secondary structure in regulating gene expression.

Here, we use a high-throughput, sequencing-based, structure-mapping approach in conjunction with three other RNA sequencing methodologies (RNA sequencing [RNA-seq], smRNA sequencing [smRNA-seq], and ribosome-bound sequencing [ribo-seq]) to determine the regulatory significance of RNA secondary structure across the *Arabidopsis* transcriptome. We use this analysis as an unbiased means to produce a comprehensive collection of RNA secondary structure models as well as reveal key structural patterns and functions of RNA folding on a transcriptome-wide scale for this model plant.

RESULTS

Using dsRNA- and Single-Stranded RNA-seq Data to Develop Experimentally Derived Models of mRNA Secondary Structure on a Genome-Wide Scale

The secondary structure of all eukaryotic mRNA molecules is dictated by specific base pairing interactions encoded within their primary nucleotide sequence (Cooper et al., 2009; Cruz and Westhof, 2009; Sharp, 2009). We have previously produced a transcriptome-wide collection of RNA secondary structure models for *Arabidopsis* using only information for paired (dsRNA) portions of RNAs (Zheng et al., 2010). However, we have recently found that having information for both paired and unpaired portions of RNAs allows for higher resolution structural models (Li et al., 2012). Therefore, we also produced data for the unpaired portion of the *Arabidopsis* transcriptome using our single-stranded RNA sequencing (ssRNA-seq) approach (Li et al., 2012; see Supplemental Figure 1 online). We then used the combination of dsRNA-seq (paired regions) (Zheng et al., 2010) and ssRNA-seq (unpaired regions) data to produce experimentally derived structural models of all *Arabidopsis* mRNAs detected in this study (see Figures 1A and 1B for examples of a highly and a lowly structured transcript, respectively) (see Methods). All of these structural models can be visualized using the genome browser at http://gregorylab.bio.upenn.edu/annoj_at9_structure/.

It is worth noting that all of the high-throughput RNA sequencing libraries produced for this study (ssRNA-seq, RNA-seq, and ribo-seq), as well as previously for dsRNA-seq (Zheng et al., 2010), were depleted for rRNA but were not subjected to a poly(A)$^+$ selection step. In this way, they lack rRNA but contain sequencing information for all other classes of RNAs.

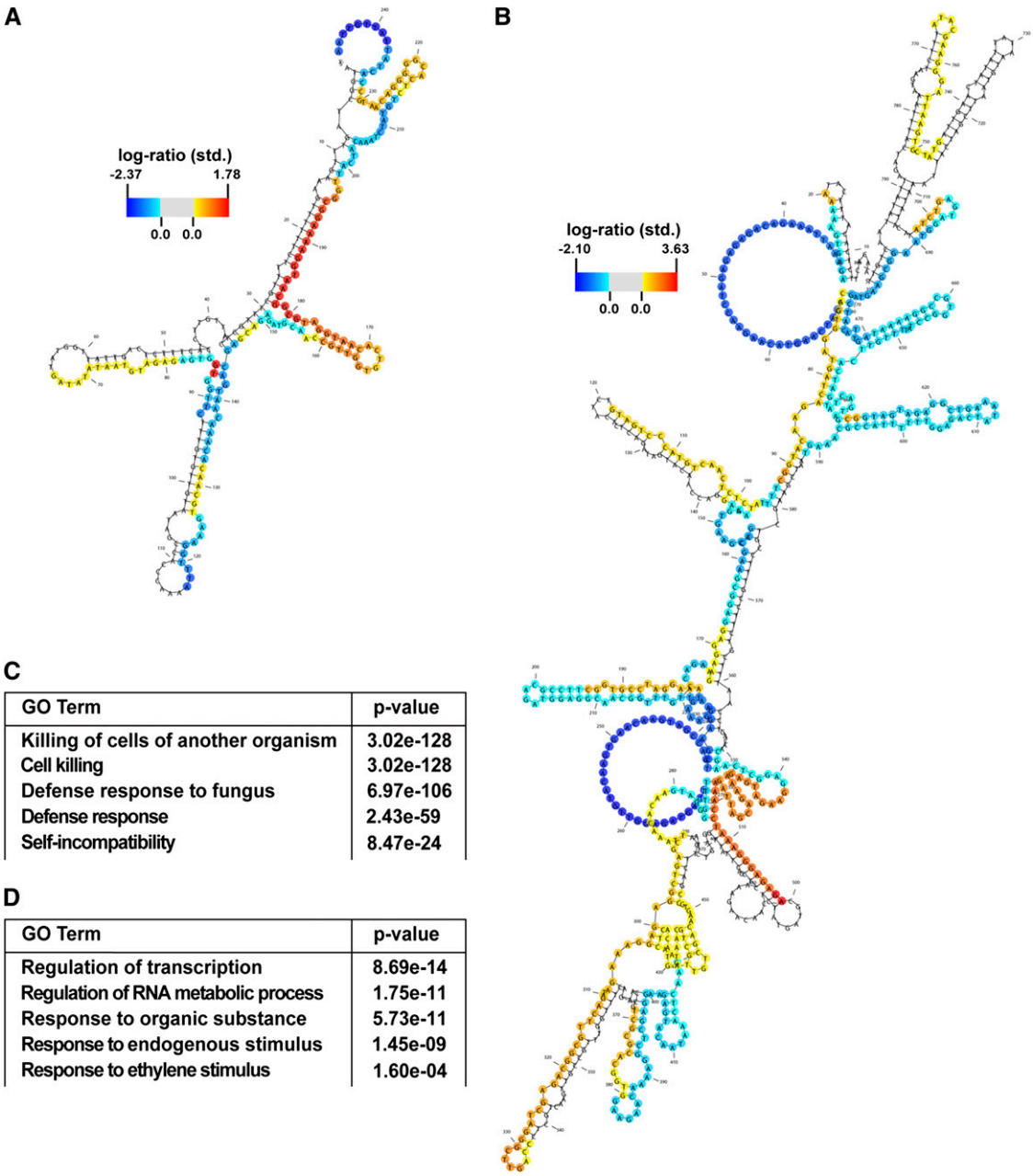
We then interrogated overrepresented biological processes for the most and least structured transcripts using Gene Ontology analysis. The four most significantly overrepresented biological processes for highly structured RNAs were related to defense response (Figures 1A and 1C). In fact, many of the highly structured mRNAs in *Arabidopsis* encode DEFENSIN-LIKE (DEFL) (Figure 1A), LOW MOLECULAR WEIGHT CYSTEINE-RICH, SCARECROW-LIKE, and PATHOGENESIS-RELATED proteins, all of which are known to function in plant pathogen responses (see Supplemental Data Set 1 online). We also found that the least structured transcripts were enriched in basic biological processes, such as "regulation of transcription" and "RNA metabolic processes" (Figures 1B and 1D). Thus, our findings suggest that pathogen-responsive transcripts tend to be the most highly folded in *Arabidopsis*, possibly due to the increased prevalence of regulatory structural elements.

Identification and Characterization of Structure Hot Spots and Cold Spots in the *Arabidopsis* Transcriptome

Next, we used a peak finding-based analysis (Muiño et al., 2011) to identify genomic regions that were significantly enriched (structure hot spots) or depleted (structure cold spots) for RNA base pairing (see Methods). This analysis identified 32,356 and 33,187 structure hot spots and cold spots, respectively (Figures 2A and 2B; see Supplemental Data Sets 2 and 3 online; data can also be downloaded from http://gregorylab.bio.upenn.edu/ArabidopsisStructure/). As expected, transposable elements were found to be a rich source of structure hot spots (Figure 2A). This is not surprising because transcriptional silencing of these regions is likely mediated by small silencing RNAs that require a self-complementary fold-back structure and/or dsRNA intermediate for their biogenesis (Willmann et al., 2011). In support of this hypothesis, we found that duplex (dsRNA molecules arising from both strands) and nonduplex (i.e., secondary structure of an RNA molecule) structure hot spots are equally likely to arise from transposable elements (see Supplemental Figures 2A and 2B online). These results reveal that transposon transcripts are substrates for dsRNA production (see Supplemental Figure 2A online) and contain highly structured regions (see Supplemental Figure 2B online) in *Arabidopsis* cells, both of which can be used by DCL proteins for smRNA production from these RNAs.

We also found numerous structure hot spots in all portions of protein-coding mRNAs (Figure 2A), while structure cold spots were almost entirely localized to gene transcripts especially within the coding sequence (CDS; protein-coding region) (Figure 2B). The enrichment of structure cold spots in protein-coding mRNAs was similar to what we observed in *Drosophila melanogaster* and *Caenorhabditis elegans* (Li et al., 2012), suggesting that large portions of eukaryotic transcripts are unpaired. This may be to facilitate ribosomal translocation during translation.

We then examined the average conservation levels (measured as average phastCons scores) for all base pairs of the structure hot spots and cold spots compared with same-sized flanking regions using a seven-way comparative genomics approach (Zheng et al., 2010; see Methods) (Figures 2C and 2D). Using this approach, we found that both hot spot and cold spot sequences were significantly more evolutionarily conserved than flanking regions (Figures 2C and 2D; P value $\rightarrow$ 0). Importantly, this was true for hot spots and cold spots in both exonic and



C

GO Term	p-value
Killing of cells of another organism	3.02e-128
Cell killing	3.02e-128
Defense response to fungus	6.97e-106
Defense response	2.43e-59
Self-incompatibility	8.47e-24

D

GO Term	p-value
Regulation of transcription	8.69e-14
Regulation of RNA metabolic process	1.75e-11
Response to organic substance	5.73e-11
Response to endogenous stimulus	1.45e-09
Response to ethylene stimulus	1.60e-04

Figure 1. Experimentally Determined Models of RNA Secondary Structure for the *Arabidopsis* Transcriptome.

(A) and **(B)** Models of secondary structure for the *Arabidopsis At1g33607.1* (highly structured, encodes a DEFL family protein) **(A)** and *HB40 (At4g36740.1)* (lowly structured, encodes a HOMEOBOX protein) **(B)** transcripts determined by our high-throughput sequencing-based, structure-mapping approach. The heat scales indicate the standardized $\log_2$ ratio of dsRNA-seq to ssRNA-seq reads (see Methods) at each base position. **(C)** and **(D)** The most significantly enriched biological processes (and corresponding P values) for the top 10% most highly **(C)** and lowly **(D)** structured *Arabidopsis* mRNAs. GO, Gene Ontology.

intronic portions of mRNAs (Figures 2C and 2D). These results suggest that significantly structured and unstructured regions of *Arabidopsis* protein-coding mRNAs have undergone purifying selection during plant evolution and are therefore likely to be functionally important. It is worth noting that our findings provide further evidence that our structure-mapping approach is identifying bona fide structural elements within protein-coding transcripts of this model plant.

We have previously observed that structure hot spots and cold spots are associated with distinct histone modifications in animals (Li et al., 2012). To determine if this association is also present in *Arabidopsis*, we examined the overlap between genome-

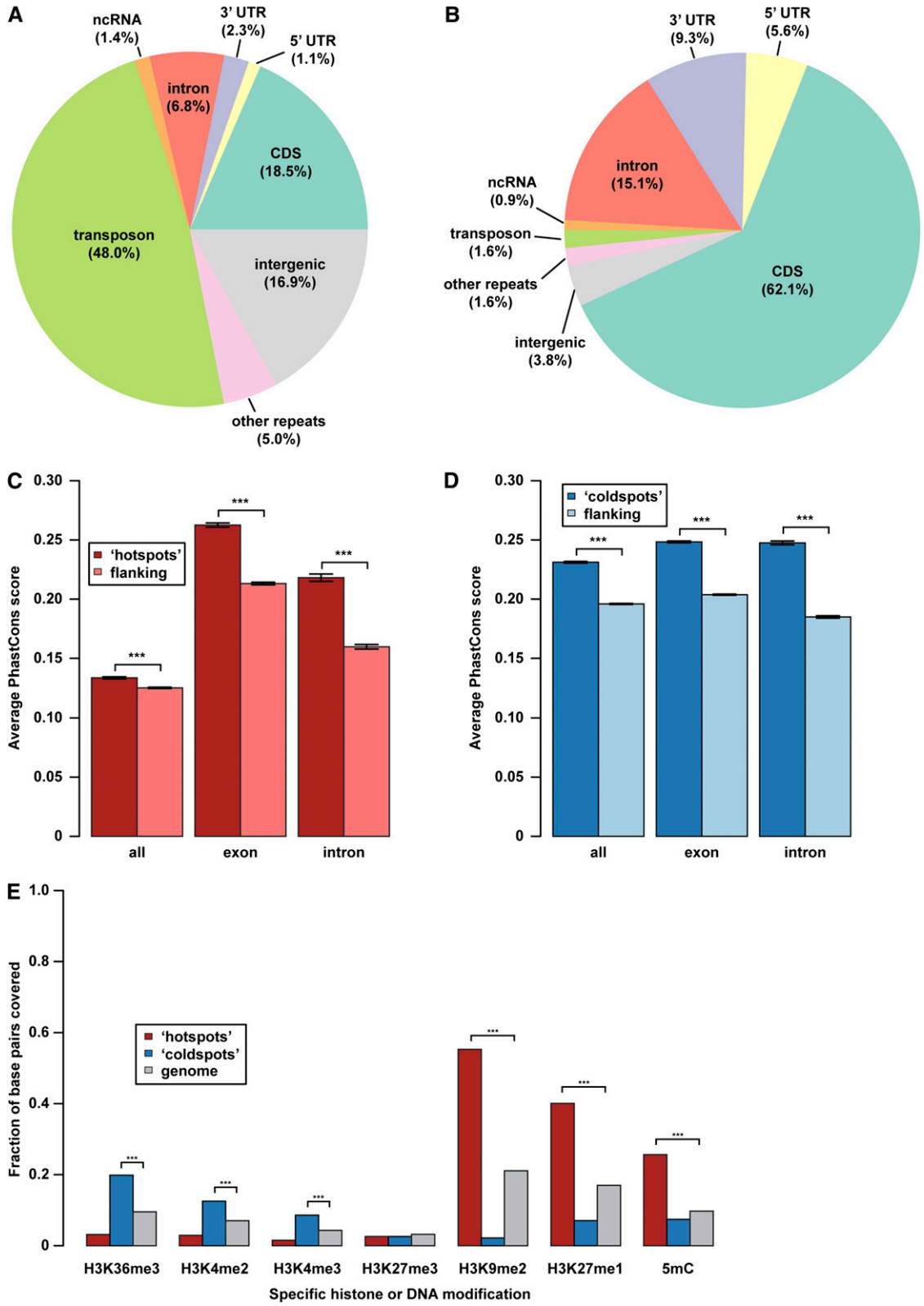


Figure 2. Identification and Characterization of Structure Hot Spots and Cold Spots in the *Arabidopsis* Transcriptome.

(A) and **(B)** Pie charts showing functional classification of structure hot spots **(A)** and cold spots **(B)**. ncRNA, noncoding RNA.

wide histone modifications determined previously (Bernatavichute et al., 2008; Jacob et al., 2010; Roudier et al., 2011) and structure hot spots or cold spots. We found that structure hot spots were significantly (all P values → 0) enriched for the siRNA-mediated, repressive, heterochromatin-associated histone 3 lysine 9 dimethylation (H3K9me2) and H3K27me1 modifications as well as DNA cytosine methylation (5mC) (Figure 2E). A closer examination of the structure hot spots that overlap these repressive modifications revealed they are distributed across duplex (dsRNA molecules arising from both strands) and nonduplex (secondary structure of an RNA molecule) RNA regions (see Supplemental Figure 2C online). In fact, the overall structural profile of the nonduplex structure hot spots that overlap with all of these siRNA-mediated, repressive, heterochromatic modifications is that of large stem regions and clusters of smaller stem loops (see Supplemental Figures 2D to 2F online). In total, these findings provide evidence for the importance of RNA secondary structure in siRNA-mediated epigenetic pathways.

Conversely, structure cold spots were significantly (all P values → 0) enriched for the activating, euchromatic H3K4me2, H3K4me3, and H3K36me3 histone modifications (Figure 2E). These results are in accordance with the overall enrichment of structure hot spots in transposons and cold spots in protein-coding transcripts (Figures 2A and 2B). In support of the accuracy of these results, the heterochromatic histone mark H3K27me3 that is not associated with siRNAs or active gene transcription was not correlated with either structure hot spots or cold spots. These findings suggest that protein-coding transcripts favor less structured conformations, possibly to evade recognition by the smRNA processing machinery of epigenetic regulatory pathways.

Secondary Structure Demarcates Translation and miRNA Target Sites in *Arabidopsis* mRNAs

To identify specific patterns within the secondary structure of *Arabidopsis* mRNAs, we examined the average structure score (see Methods) across the coding region (CDS) and both 5′ and 3′ untranslated regions (UTRs) of all detected protein-coding transcripts (Figure 3A). We identified significant (P value → 0) decreases in average structure score near the start and stop codons of the CDS, revealing increased accessibility of the RNA at the regions where protein translation begins and ends (Figure 3A). We and others also observed this same trend for yeast, *Drosophila*, and *C. elegans* mRNAs (Kozak, 2005; Kertesz et al., 2010; Li et al., 2012), strongly indicating that it may be a general feature of eukaryotic protein-coding transcripts. Overall, we also observed that both UTRs were, on average, less structured than the coding region in *Arabidopsis* mRNAs, which was also observed for yeast transcripts (Kertesz et al., 2010). Taken together, these results reveal that there are conserved structural patterns within eukaryotic protein-coding mRNAs and suggest that these features may affect translation.

Given the ability to interrogate average secondary structure across protein-coding transcripts, we next examined the site-specific average structure scores at positions within predicted miRNA target sites (Dai and Zhao, 2011) and the 50 base pairs of sequence up- and downstream of these regions (Figure 3B). This analysis revealed significantly (P value → 0) decreased base pairing across the entire length of miRNA binding sites compared with the flanking regions in target mRNAs (Figure 3C). These findings are similar to what was previously hypothesized for *Arabidopsis* miRNA target sites based on energy-derived RNA secondary structure predictions (Gu et al., 2012) and what we have formerly observed in *C. elegans* (Li et al., 2012). Therefore, our results provide strong experimental support for this predicted structural trend in *Arabidopsis* miRNA binding sites. Together, these findings reveal that miRNA binding sites are less structured than their flanking regions, likely to allow more efficient interaction between a miRNA and its target mRNAs as previously suggested (Brennecke et al., 2005; Kertesz et al., 2007; Long et al., 2007; Dai et al., 2011).

RNA Secondary Structure Regulates Overall Transcript Abundance in *Arabidopsis*

RNA secondary structure is known to regulate the levels of some mRNAs in eukaryotes (Goodarzi et al., 2012). However, the global effect of this feature on the steady state abundance of *Arabidopsis* mRNAs is almost entirely unknown. Therefore, we determined the effect of increasing RNA secondary structure on the abundance of each mRNA detectable by RNA-seq (see Methods) and found that RNA folding had a significant negative effect (Pearson correlation $r = -0.45$, P value → 0) on total transcript levels (Figure 4A; see Supplemental Figure 3A online). In fact, our findings indicate that RNAs with low levels of secondary structure will be on average more abundant in the transcriptome and vice versa (see Supplemental Figure 3A and Supplemental Table 1 online). We confirmed this observation using quantitative RT-PCR (qRT-PCR) on five highly (top 10%; see Supplemental Figure 3A online) and seven lowly (bottom 10%; see Supplemental Figure 3A online) structured mRNAs (12 total mRNAs). From this analysis, we found that the less structured mRNAs were all significantly (P value < 0.001) more abundant than those transcripts with high levels of folding (Figure 4B; see Supplemental Figure 3B online). In total, these findings reveal that significant levels of mRNA secondary structure have a strong negative regulatory effect on the overall abundance of mRNAs in the *Arabidopsis* transcriptome.

Figure 2. (continued).

(C) and **(D)** Conservation scores for structure hot spots **(C)** or cold spots **(D)** (dark-red and blue, respectively) versus flanking regions. Higher values indicate more conservation within the seven plant species (see Methods). ***P value → 0. P values were calculated by a *t* test.

(E) The fraction of base pairs within structure hot spots (red), cold spots (blue), and the entire genome (gray) that are marked by specific histone modifications (as indicated in the figure). Values are given as the fraction of all base positions for hot spots, cold spots, or the entire genome (control) that are associated with the given epigenetic mark. ***P value → 0. P values were calculated using a χ^2 test.

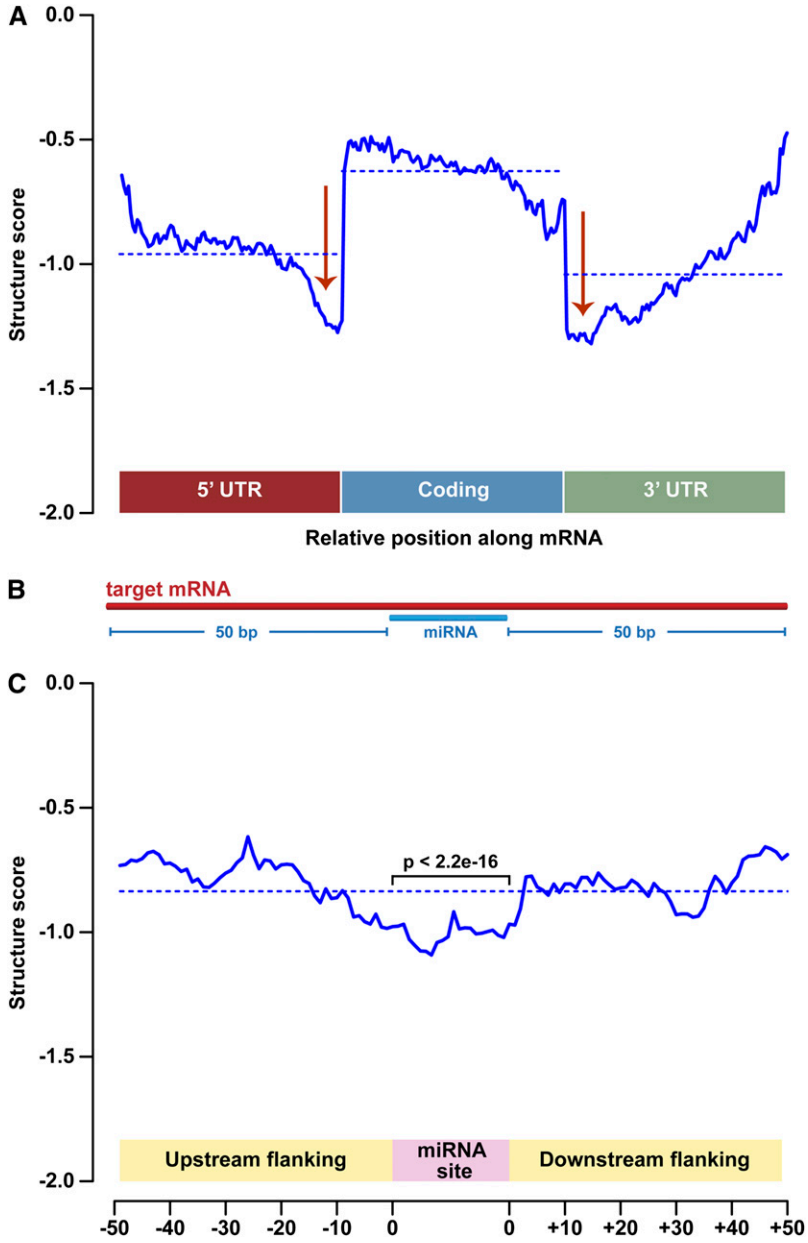


Figure 3. Secondary Structure Marks Translation Initiation and Termination as well as miRNA Target Sites in *Arabidopsis* mRNAs.

(A) The average structure score plotted over the 5′ UTR, CDS, and 3′ UTR of all detectable protein-coding transcripts for *Arabidopsis*. The overall average for each specific transcript region is shown as a dotted line. Red arrows highlight significant (P value → 0, *t* test) dips in secondary structure that occur at the junctions between the UTRs and the coding region.

(B) Model depicting our analysis of RNA secondary structure at miRNA binding sites in target mRNAs.

(C) The average structure score across miRNA binding sites and for 50 bp up- and downstream flanking regions in *Arabidopsis* target transcripts. The overall structure score average for the entire ~121-bp region is shown as a dotted line. P value was calculated by a *t* test.

We wondered if mRNA degradation and/or smRNA processing could explain the relationship between secondary structure and overall transcript abundance. To test this, we normalized two different sets of previously published genome-wide RNA degradation (degradome) data (Addo-Quaye et al., 2008; Gregory et al., 2008) by total transcript abundance as measured by RNA-seq data (see Methods) to ascertain the levels of degradation for every detectable mRNA (Figure 5A; see Supplemental Figure 4 online). We found a significant positive correlation (Pearson correlation $r = \sim 0.21$, P → 0) between the overall structure score and degradation level of *Arabidopsis* mRNAs (Figure 5A; see Supplemental Figure 4 and Supplemental Data Set 1 online), indicating that highly folded mRNAs tend to be degraded more frequently than less structured transcripts. Interestingly, the tendency

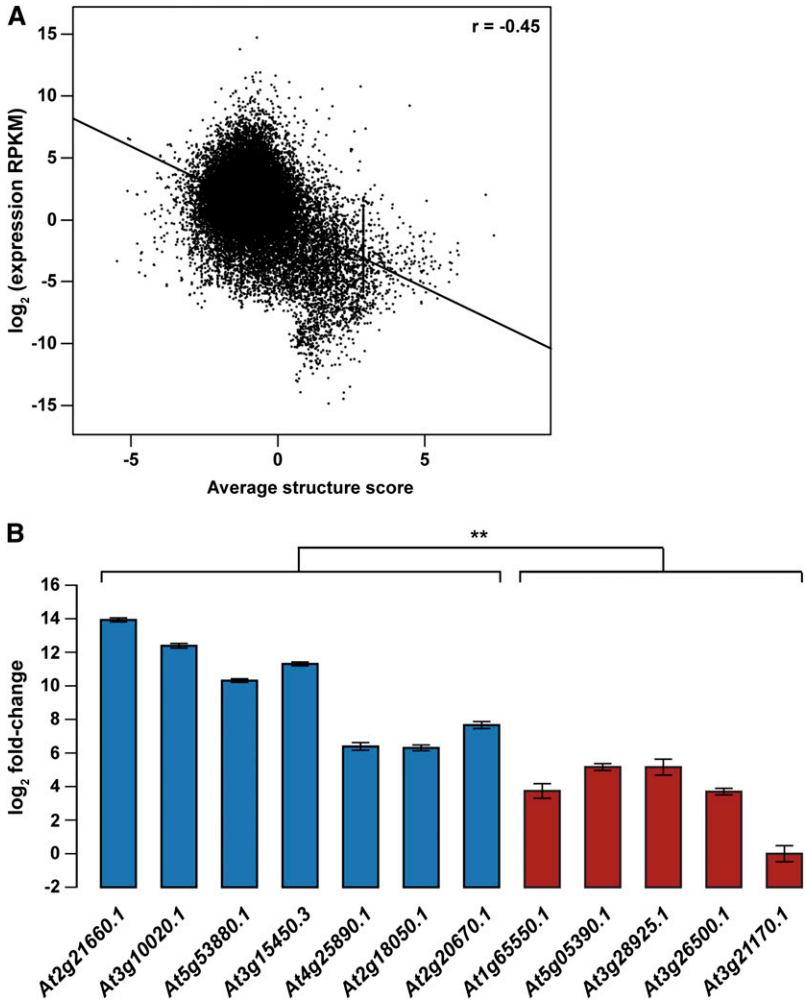


Figure 4. RNA Secondary Structure Regulates the Overall Abundance of *Arabidopsis* mRNAs.

(A) The average structure score (*x* axis) is plotted against average expression values determined by RNA-seq (*y* axis) for all detectable *Arabidopsis* mRNAs.

(B) Random hexamer-primed qRT-PCR analysis of seven lowly (blue bars) and five highly (red bars) structured *Arabidopsis* mRNAs. Error bars, ± SE. **P value < 0.001. P value was calculated by a one-tailed *t* test.

for structure and transcript degradation level to be significantly correlated is especially obvious for mRNAs with predicted miRNA target sites (see Supplemental Figures 4A to 4C online; Pearson correlation $r = 0.36$, P value → 0). This is because highly structured RNAs are often targeted for degradation both by miRNA binding events and intrinsic structural features (see Supplemental Figures 4A to 4C and 4E and Supplemental Data Set 1 online). Taken together, these results suggest that RNA secondary structure is an intrinsically destabilizing feature of protein-coding mRNAs in *Arabidopsis*.

The high levels of degradation observed for highly structured mRNAs could also be a consequence of increased smRNA processing from these transcripts. Therefore, we used smRNA-seq data to determine if increasing mRNA secondary structure had a significant effect on the total levels of smRNAs processed from mRNAs (see Methods). Using this approach, we found a significant positive correlation (Pearson correlation $r = 0.62$, P value → 0) between increasing mRNA secondary structure and higher levels of sense smRNA production (Figure 5B; see Supplemental Figure 5A and Supplemental Data Set 1 online). As expected, the regions of mRNAs that are processed into smRNAs are significantly (P value → 0) more structured than the regions that are not cleaved into smRNAs (Figure 5C). In fact, the overall structural profile of these regions is that of large stem regions and clusters of smaller stem loops (see Supplemental Figures 5D to 5G online). Interestingly, we also found a similar trend (Pearson correlation $r = 0.65$) for production of smRNAs mapping to the antisense strand (see Supplemental Figures 5B and 5C and Supplemental Data Set 1 online). Together, these findings suggest that mRNA secondary structure results in initial processing of the sense transcript, which then becomes a target for dsRNA synthesis (likely by an RNA-dependent RNA polymerase)

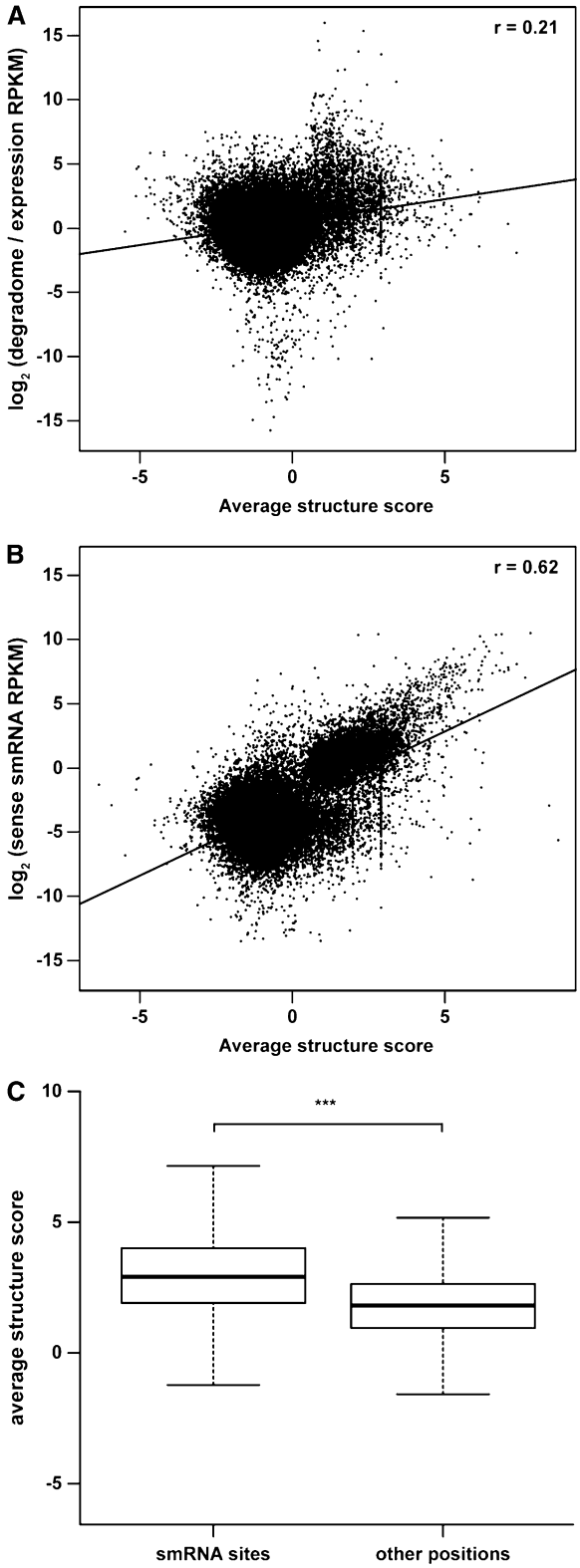


Figure 5. RNA Secondary Structure Promotes the Degradation and Processing of *Arabidopsis* mRNAs into smRNAs.

(A) The average structure score (*x* axis) is plotted against average degradation values determined by correcting degradome (Gregory et al., 2008) values by RNA-seq (*y* axis) for all detectable *Arabidopsis* mRNAs.

(B) The average structure score (*x* axis) is plotted against the total abundance of smRNAs present per transcript in the sense orientation as determined by smRNA-seq (*y* axis) for all detectable *Arabidopsis* mRNAs.

(C) The average structure score (*y* axis) of mRNA regions processed into smRNAs (left box, smRNA sites) compared with those that are not (right box, other positions). ***P value → 0. P value was calculated by *t* test.

and subsequent production of both sense and antisense smRNAs. In total, our results reveal that highly folded mRNAs tend to be less abundant in the *Arabidopsis* transcriptome because they are more often degraded and/or processed into smRNAs than transcripts that are less structured.

Translation Is Affected by the Folding of *Arabidopsis* mRNAs

The significant effect of RNA structure on mRNA abundance (Figure 4) as well as the dips in structure at the beginning and end of the CDS of most *Arabidopsis* transcripts (Figure 3) suggested that this feature may also regulate protein translation. Therefore, we tested the effect of RNA folding on total ribosome association, which was determined by normalizing ribo-seq data (see Methods) by total transcript abundance as measured by RNA-seq (see Methods) for every detectable mRNA. Using this approach, we found that RNA folding had a significant positive effect (Pearson correlation $r = 0.37$, P value → 0) on overall ribosome association (Figure 6A; see Supplemental Figure 6A and Supplemental Data Set 1 online). In fact, our findings reveal that RNAs with high levels of secondary structure will be on average more ribosome bound than less structured transcripts (see Supplemental Figure 6A and Supplemental Data Set 1 online). We confirmed this observation using qRT-PCR on ribosome-associated RNA fractions (see Methods) for five highly (top 10%; see Supplemental Figure 6A online) and seven lowly (bottom 10%, see Supplemental Figure 6A online) structured mRNAs (12 total mRNAs) and normalizing these values by total mRNA abundance as measured by qRT-PCR of total RNA. From this analysis, we found that the more structured transcripts were significantly (P value < 0.001) more ribosome-associated than the less structured transcripts (Figure 6B; see Supplemental Figure 6B online). In total, these findings reveal that mRNA secondary structure exerts a significant effect on translation of mRNAs into proteins and suggest that mRNA folding regulates gene expression at multiple levels in *Arabidopsis*.

DISCUSSION

Here, we report a characterization of RNA secondary structure and its regulatory significance across a plant transcriptome. To improve our initial structure predictions (Zheng et al., 2010), we have now produced data for the unpaired portion of the *Arabidopsis* transcriptome (see Supplemental Figure 1 online). The combination of dsRNA-seq (Zheng et al., 2010) and ssRNA-seq (see Supplemental Figure 1 online) data allowed us to produce a higher resolution view of RNA secondary structure in this model plant (Figure 1).

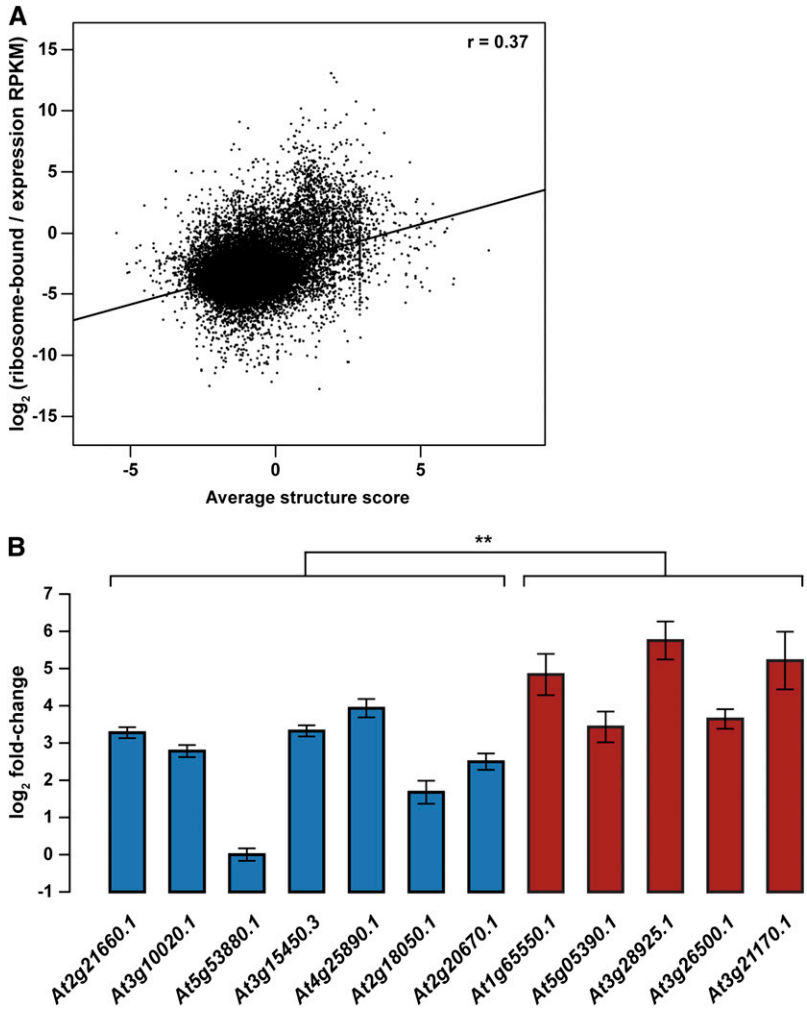


Figure 6. RNA Secondary Structure Regulates the Ribosome Association of *Arabidopsis* mRNAs.

(A) The average structure score (*x* axis) is plotted against average ribosome association values determined by normalizing ribo-seq values by RNA-seq (*y* axis) for all detectable *Arabidopsis* mRNAs.

(B) Random hexamer-primed qRT-PCR analysis of seven lowly (blue bars) and five highly (red bars) structured *Arabidopsis* mRNAs using ribosome-bound RNA fractions with values corrected by total RNA abundance as also measured by qRT-PCR. Error bars indicate ±SE. **P value < 0.001. P value was calculated by a one-tailed *t* test.

By combining these two data sets, we were also able to identify structure hot spots and cold spots genome wide (Figures 2A and 2B). Interestingly, we found significant correlations between RNA structure hot spots and genomic sites of siRNA-mediated, heterochromatic histone modifications and DNA methylation (Figure 2E; see Supplemental Figure 2 online), providing experimental support for the importance of RNA secondary structure in epigenetic regulatory pathways. Additionally, it is well accepted that DCL1 recognizes structured regions of RNA molecules for production of smRNAs (Voinnet, 2009; Willmann et al., 2011), and our transcriptome-wide structural studies also support this model of substrate recognition for other DCL proteins (e.g., DCL3) in *Arabidopsis* (see Supplemental Figure 2 online).

We also identified a strong correlation between structure cold spots and activating, euchromatic histone modifications (Figure 2E), suggesting that actively transcribed mRNAs prefer highly unpaired secondary conformations. This result fits with our observation that RNA secondary structure has a significant anti-correlation with overall transcript abundance (Figure 4). In total, our findings suggest that *Arabidopsis* protein-coding mRNAs are less structured than other classes of RNAs, in general, and this feature is central to their proper regulation (more on this below).

Interestingly, we also found that many of our identified structure hot spots and cold spots in protein-coding mRNAs are evolutionarily more conserved than their flanking sequences and, therefore, likely have some sort of function outside of their capacity to encode protein sequence (Figures 2C and 2D). Future experiments will be aimed at addressing the specific functions of such mRNA regions within the *Arabidopsis* transcriptome.

Structural Patterns in *Arabidopsis* mRNAs

Given the ability to globally characterize RNA secondary structure, we examined the average folding patterns of protein-coding transcripts. This analysis uncovered specific mRNA structural features (Figure 3). For example, we revealed a significant decrease in mRNA secondary structure near both the start and stop codons of the CDS. We also observed these same structural features in *Drosophila* and *C. elegans* protein-coding transcripts (Li et al., 2012). Together, these findings indicate that specific folding patterns demarcate the protein-coding region of eukaryotic mRNAs and suggest that secondary structure has a regulatory effect on protein translation (Figures 3A and 6).

Additionally, we found that on average both the 5′ and 3′ UTRs are less structured than the coding region of *Arabidopsis* transcripts (Figure 3A). This is similar to what was observed for yeast transcripts (Kertesz et al., 2010) but opposite of what we have found for *Drosophila* and *C. elegans* mRNAs (Li et al., 2012). These results suggest that there is not a general pattern of RNA secondary structure that encompasses all eukaryotic protein-coding transcriptomes. Therefore, future genome-wide, structure-mapping projects will be necessary to define RNA folding patterns in other species.

Recently, it was hypothesized that specific synonymous codons are selected around miRNA binding sites in plants to decrease base pairing interactions in these regions of target transcripts (Gu et al., 2012). We tested this hypothesis using our experimentally determined structure data for *Arabidopsis* and found that miRNA binding sites are on average significantly less structured than both 5′ and 3′ flanking regions (Figures 3B and 3C). These results revealed that the sites targeted by miRNAs are more accessible to interaction with miRNA-loaded RNA-induced silencing complex RISC. The combination of our results and those of the previous study on four plants (*Arabidopsis*, rice [*Oryza sativa*], *Populus trichocarpa*, and maize [*Zea mays*]) (Gu et al., 2012) suggest that decreased structure over miRNA binding sites is a common feature within plant transcriptomes.

Gene Expression Regulation by RNA Secondary Structure in *Arabidopsis*

We also observed that a high degree of RNA secondary structure has a significant negative effect on overall transcript abundance (Figure 4; see Supplemental Figure 3 online). In fact, we found that the 10% most highly structured transcripts were found at significantly lower levels on average compared with the least folded *Arabidopsis* mRNAs (see Supplemental Figure 3A online). This was somewhat surprising because RNA binding proteins tend to bind highly structured regions of mRNAs and often have a stabilizing effect on bound transcripts (Keene, 2001; Glisovic et al., 2008; Khalil and Rinn, 2011; Konig et al., 2012). However, our data suggest that highly structured transcripts are often tightly regulated, pathogen-responsive genes in *Arabidopsis* (Figure 1C). We hypothesize that the cell can recognize these response-specific transcripts based on their secondary structure. Thus, once the infection is over, the highly structured, inducible mRNAs can then be quickly recognized and removed, often by the RNA degradation and/or silencing machinery (Figure 5; see Supplemental Figures 4 and 5 online).

By interrogating the effects of increasing structure on RNA stability and smRNA processing from *Arabidopsis* protein-coding RNAs, we found that highly structured RNAs are more highly degraded and/or processed into smRNAs compared with less folded transcripts (Figure 5; see Supplemental Figures 4 and 5 online). For instance, the regions of mRNAs processed into smRNAs are more structured than the parts that are not cleaved into smRNAs (Figure 5C). In fact, these regions often form stem loops that are reminiscent of miRNA-like structures (see Supplemental Figures 5D to 5G online). These results provide an explanation for why highly structured transcripts are on average less abundant (Figure 4; see Supplemental Figure 3 online) and are intriguing given our observations that RNA secondary structure likely results in the biogenesis of numerous classes of smRNAs (Figure 5; see Supplemental Figures 2 and 4 online). Interestingly, our findings suggest that *Arabidopsis* protein-coding mRNAs on average maintain lower levels of RNA secondary structure than noncoding transcripts (Figures 2 and 3) likely to evade recognition by the smRNA processing machinery of RNA silencing pathways. In fact, these results suggest that RNA secondary structure is potentially the feature that distinguishes endogenous RNAs (low structure) from foreign RNAs (e.g., viruses and transposons; high structure). In support of this hypothesis, structure cold spots are highly enriched in *Arabidopsis* protein-coding mRNAs, while hot spots are commonly found in transposable elements (Figures 2A and 2B). Additionally, a structural analysis of the HIV-1 RNA genome has demonstrated that this viral RNA is very highly structured (Watts et al., 2009). In total, our global view of RNA secondary structure provides evidence that this feature may allow the RNA silencing machinery of plants to differentiate endogenous from foreign RNA molecules.

Finally, our analyses revealed that highly structured RNAs tend to be more ribosome-bound on a per transcript basis (Figure 6; see Supplemental Figure 6 online). A number of reasons could explain these observations. For example, RNA secondary structure could decrease the efficiency of translation initiation and/or termination, as suggested by the average structural profile of *Arabidopsis* mRNAs (Figure 3A). Additionally, RNA folding could cause ribosomes to cluster on mRNAs due to translocation slowing and/or pausing. We favor the hypothesis that RNA secondary structure increases ribosome occupancy by both of these models and probably others not discussed here, with each different structural element within protein-coding mRNAs affecting ribosomes uniquely. It will be important to explore the consequences of ribosome–RNA structure interactions further in future experiments, as understanding how each of these interactions affects the efficiency and speed of protein translation will likely be important for the future of crop improvement and plant-focused biofuel research. In total, our study has uncovered that RNA secondary structure has pleiotropic regulatory effects on *Arabidopsis* gene expression.

METHODS

Plant Materials

Immature flower bud clusters from the Columbia-0 ecotype of *Arabidopsis thaliana* grown under 16-h days were used for all experiments and analyses described in this study.

RNA Analyses

For all experiments performed herein, RNA was isolated using the miRNeasy mini kit (Qiagen). Random hexamer-primed cDNA was made for at least three biological replicates per experiment. Transcripts were then quantified by qRT-PCR using the comparative threshold cycle method ($\Delta\Delta C_t$), using *Actin 2* (*At3g18780*) as the endogenous reference and the lowest expressed transcript for renormalization. Primers used for qRT-PCR are listed in Supplemental Table 1 online.

Sequencing Library Preparation

The ssRNA-seq library (Li et al., 2012), as well as the smRNA-seq library (Zheng et al., 2010), were constructed as previously described. The RNA-seq libraries were produced using the SOLiD Total RNA-seq library preparation kit (Applied Biosystems).

Ribo-seq libraries were made using ribosome-associated mRNAs from unopened flower buds that were isolated by differential centrifugation according to Mustroph et al. (2009) with the following modifications. The ribosomes and associated mRNAs pelleted by centrifugation through a Suc cushion were resuspended in 0.2 M Tris, pH 8.0, 0.2 M KCl, 0.035 M $MgCl_2$, 50 μg/mL chloramphenicol, and 50 μg/mL cycloheximide. Forty micrograms of resuspended RNA was centrifuged over a 15 to 60% Suc gradient (0.04 M Tris, pH 8.0, 0.02 M KCl, 0.02 $MgCl_2$, 5 μg/mL chloramphenicol, and 5 μg/mL cycloheximide). Following centrifugation, 50 μL fractions of the gradient were isolated and the OD_{260} of each was measured. The monosomal and polysomal fractions were pooled, and the RNA was isolated using the Qiagen miRNeasy mini kit. Eight micrograms of isolated RNA was depleted of rRNA using the RiboMinus plant kit (Life Technologies), fragmented using RNA fragmentation reagents (Ambion), treated with T4 PNK (NEB) to repair 5′ and 3′ ends, and used for library preparation using the Illumina TruSeq smRNA-seq library preparation kit and accompanying protocols (Illumina).

High-Throughput Sequencing

The smRNA-seq library was sequenced on an Illumina GA2 analyzer, the ribo-seq and ssRNA-seq libraries were sequenced on an Illumina Hi-Seq2000 platform, and two biological replicates of RNA-seq libraries were sequenced on an ABI SOLiD 3+ (Applied Biosystems). All sequencing was performed according to the manufacturer's instructions. Detailed information for each library is summarized in Supplemental Table 2 online.

Read Processing and Alignment

Read processing and alignment were performed as previously described (Zheng et al., 2010; Li et al., 2012) with slight modifications. Because our various sequencing libraries were prepared and sequenced separately with a significant time span, reads from these libraries were processed and aligned by several different variants of a balanced pipeline as previously described (Zheng et al., 2010; Li et al., 2012) each with the following modifications.

Trimming of 3′-Adapters

All raw reads (except ribo-seq reads) were first trimmed to remove 3′ sequencing adapters using VectorStrip (from EBI EMBOSS package) or Cutadapt (Martin, 2011). Reads that were not trimmed were kept unchanged.

Reducing to Nonredundant Tags

Reads after trimming were reduced to nonredundant tags to save processing time and space for subsequent steps. The clone abundance of each nonredundant tag was recorded and used in all subsequent abundance calculations.

Mapping to Reference Arabidopsis Genome

All reads were mapped to the *Arabidopsis* genome (TAIR9 assembly) using cross-match (from Phred/Phrap package) or Bowtie (v0.12.7) (Langmead et al., 2009). The running parameters of these programs were carefully tuned (see Supplemental Table 2 online). A subsequent parsing step was implemented to enforce the mismatch standards (see Supplemental Table 2 online), as well as to determine the true insert length for untrimmed reads. For Bowtie alignments, an additional seed mismatch percentage restriction was required to control mismatch percentage in the high-quality 5′ end of sequencing reads. Multiply mapped reads were allowed for all of these libraries. For RNA-seq reads in SOLiD color space, both the reference *Arabidopsis* genome and raw reads were first converted into pseudo-base space prior to the mapping using the MAQ program (http://maq.sourceforge.net/), and the mapping coordinates were then restored into normal sequence space. It is of note that the plus and minus strands of the genome were converted into pseudo-base space independently. This is because the sequences of the two strands in pseudo-base space are dependent on their respective first bases and thus are not complementary to each other.

Summary of Mapped Reads

For reads with multiple hits in the genome (multiply mapped), a max-diverge filter was also implemented, to select only those hits with a mismatch percentage no more than a fixed max-diverge percentage to the best hit of each read. This filter is similar to UCSC BLAT and Bowtie "–best" mode. Mapped reads were then summarized to retain all pertinent information, such as their length, clone abundance, and number of locations in the genome. All of this information was loaded into local MySQL databases for further queries, such as determining the transcript abundance in various libraries used in this study (see directly below).

Determining Protein-Coding mRNA Abundance of Various Libraries

To determine the expression abundance of protein-coding mRNAs, all *Arabidopsis* mRNAs were searched for exon-mapped reads using the MySQL databases described above, allowing calculation of a reads per kilobase per million mapped reads (RPKM)-normalized expression value based on total mapped reads. Only sense strand-mapped reads were considered in this analysis. However, for the smRNA-seq library the antisense expression was also calculated independently.

Calculation of Per-Nucleotide Structure Scores and Use in Structure Modeling

The structure score at each base position was calculated as the generalized log ratio (glog) (Durbin et al., 2002; Huber et al., 2002) of dsRNA-seq to ssRNA-seq coverage (n_{ds}, n_{ss}) after normalization by the total number of mapped reads in each library (N_{ds} and N_{ss}), as follows:

$$S_i = \mathrm{glog}(ds_i) - \mathrm{glog}(ss_i) = \log_2\left(ds_i + \sqrt{1 + ds_i^2}\right) - \log_2\left(ss_i + \sqrt{ss_i^2}\right)$$

$$ds_i = n_{ds} * \frac{\max(N_{ds}, N_{ss})}{N_{ds}}, \quad ss_i = n_{ss} * \frac{\max(N_{ds}, N_{ss})}{N_{ss}}$$

where S_i is the structure score and ds_i and ss_i are the normalized dsRNA-seq and ssRNA-seq read coverages at position i. The generalized log ratio can tolerate 0 values (positions with no dsRNA or ssRNA read coverage) as well as being asymptotically equivalent to the standard log ratio when the coverage values are large (Durbin et al., 2002; Huber et al., 2002).

A standardized version of the structure score Z_i was used to constrain RNAfold (from the Vienna package) (Hofacker, 2003) predictions of secondary structure for each transcript:

$$Z_i = \frac{S_i - \bar{S}}{s^2}$$

where $\bar{S}$ and s^2 are the mean and SD of scores S_i for a given transcript. To determine thresholds to call paired and unpaired positions, a null distribution of standardized structure scores was calculated by randomly shuffling dsRNA and ssRNA reads and recomputing the standardized scores. Thresholds of +2.25 and −1.48 were set to call bases as paired or unpaired, respectively, based on a 5% false discovery rate. Constrained structure modeling was then performed as previously described (Li et al., 2012).

Gene Ontology Analyses

Gene Ontology analyses of enriched processes in the most and least structured protein-coding mRNAs were performed using the DAVID online tool (Huang et al., 2009a, 2009b).

Identification of Structure Hot Spots and Cold Spots

Structure hot spots and cold spots were identified using a modified version of the CSAR software package (Muiño et al., 2011). Specifically, structure scores were calculated for each base position in the genome, and regions with significantly higher or lower than background scores at an FDR of 5% were called as hot spots and cold spots, respectively. The background distribution for determining the FDR was calculated by randomly shuffling dsRNA and ssRNA reads and then computing structure scores as described above.

Conservation (phastCons) Analysis of Structure Hot Spots and Cold Spots and Their Flanking Regions

This analysis was performed using our seven-way plant comparative genomics approach as previously described (Zheng et al., 2010). Briefly, six plant genomes (*Glycine max* [Glyma1], *Populus trichocarpa* [v1.0], *Sorghum bicolor* [v1.0], *Medicago truncatula* [Mt3], *Oryza sativa* [release 6.1], and *Vitis vinifera* [Genoscape 8.4]) were aligned to the *Arabidopsis* (TAIR9) genome using the University of California Santa Cruz (UCSC) whole-genome alignment pipeline (Margulies et al., 2003). This flowering plant phastCons score data that we generated for our analyses are available for download at http://gregorylab.bio.upenn.edu/Arabidopsis Structure/.

Analysis of Secondary Structure at miRNA Target Sites

Putative miRNA target sites were downloaded from the psRNATarget Web server (Dai and Zhao, 2011), using the 243 published *Arabidopsis* miRNAs from miRBase release 16 and all protein-coding mRNA transcripts from TAIR9.

Statistical Analyses

All statistical analyses were performed using the R software package (http://www.r-project.org/), including P values for all correlation analyses. See figure legends for specific statistical tests used to assess significance.

Accession Numbers

All ssRNA- and ribo-seq data generated for this study were deposited in the Gene Expression Omnibus (GEO) under accession number GSE40209. All smRNA- and RNA-seq data used in this study were deposited in GEO under accession number GSE28524. Previously published dsRNA-seq and degradome data also used in this study can be found in GEO under the accession numbers GSE23439 and GSE11070, respectively.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Characterization of Total ssRNA-seq Reads.

Supplemental Figure 2. Comparison of Duplex versus Nonduplex Structure Hot Spots.

Supplemental Figure 3. RNA Secondary Structure Regulates the Overall Abundance of *Arabidopsis* mRNAs.

Supplemental Figure 4. Transcripts Containing Predicted miRNA Target Sites Show Significant Positive Correlation between Secondary Structure and Degradation Levels.

Supplemental Figure 5. RNA Secondary Structure Promotes the Processing of *Arabidopsis* mRNAs into smRNAs.

Supplemental Figure 6. RNA Secondary Structure Regulates the Ribosome Association of *Arabidopsis* mRNAs.

Supplemental Table 1. Oligomers Used in This Study.

Supplemental Table 2. A Brief Summary of Read Processing and Alignment Procedures Used for All High-Throughput Sequencing Libraries in This Study.

Supplemental Data Set 1. Structure Score, Expression (as RPKM Values), Degradation, smRNA Processing, and Ribosome Association Information for All *Arabidopsis* Transcripts Detected in This Study.

Supplemental Data Set 2. Information for All Structure Hot Spots Identified in This Study.

Supplemental Data Set 3. Information for All Structure Cold Spots Identified in This Study.

ACKNOWLEDGMENTS

We thank Isabelle Dragomir and Jamie Yang for technical support. This work was funded by National Science Foundation Career Award MCB-1053846 to B.D.G., National Human Genome Research Institute 5T32HG000046-13 to F.L., and National Institute of General Medical Sciences 5T32GM007229-37 to L.E.V.

AUTHOR CONTRIBUTIONS

F.L., Q.Z., and B.D.G. conceived and designed the study. F.L., L.E.V., M.R.W., and B.D.G. performed all experiments. F.L., Q.Z., L.E.V., Y.C., and B.D.G. performed all analyses. F.L., Q.Z., L.E.V., M.R.W., and B.D.G. interpreted the data. F.L., Q.Z., L.E.V., M.R.W., and B.D.G. wrote the article.

Received August 18, 2012; revised October 22, 2012; accepted October 25, 2012; published November 13, 2012.

REFERENCES

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. Curr. Biol. **18:** 758–762.

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol. **15:** 251–258.

Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell **116:** 281–297.

Baulcombe, D. (2004). RNA silencing in plants. Nature **431:** 356–363.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PLoS ONE **3:** e3156.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol. **3:** e85.

Brierley, I., Pennell, S., and Gilbert, R.J. (2007). Viral RNA pseudoknots: Versatile motifs in gene expression and replication. Nat. Rev. Microbiol. **5:** 598–610.

Buratti, E., Muro, A.F., Giombi, M., Gherbassi, D., Iaconcig, A., and Baralle, F.E. (2004). RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. Mol. Cell. Biol. **24:** 1387–1400.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell **136:** 642–655.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell **136:** 777–793.

Cruz, J.A., and Westhof, E. (2009). The dynamic landscapes of RNA architecture. Cell **136:** 604–609.

Dai, X., and Zhao, P.X. (2011). psRNATarget: A plant small RNA target analysis server. Nucleic Acids Res. **39**(Web Server issue)**:** W155–W159.

Dai, X., Zhuang, Z., and Zhao, P.X. (2011). Computational analysis of miRNA targets in plants: current status and challenges. Brief. Bioinform. **12:** 115–121.

Durbin, B.P., Hardin, J.S., Hawkins, D.M., and Rocke, D.M. (2002). A variance-stabilizing transformation for gene-expression microarray data. Bioinformatics **18**(suppl. 1)**:** S105–S110.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. **582:** 1977–1986.

Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M., and Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature **485:** 264–268.

Gregory, B.D., O'Malley, R.C., Lister, R., Urich, M.A., Tonti-Filippini, J., Chen, H., Millar, A.H., and Ecker, J.R. (2008). A link between RNA metabolism and silencing affecting *Arabidopsis* development. Dev. Cell **14:** 854–866.

Gu, W., Wang, X., Zhai, C., Xie, X., and Zhou, T. (2012). Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. Mol. Biol. Evol. **29:** 3037–3044.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. **31:** 3429–3431.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. **37:** 1–13.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4:** 44–57.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics **18** (suppl. 1)**:** S96–S104.

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature **466:** 987–991.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu. Rev. Plant Biol. **57:** 19–53.

Keene, J.D. (2001). Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. Proc. Natl. Acad. Sci. USA **98:** 7018–7024.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. **39:** 1278–1284.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H. Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. Nature **467:** 103–107.

Khalil, A.M., and Rinn, J.L. (2011). RNA-protein interactions in human health and disease. Semin. Cell Dev. Biol. **22:** 359–365.

Klasens, B.I., Das, A.T., and Berkhout, B. (1998). Inhibition of polyadenylation by stable RNA secondary structure. Nucleic Acids Res. **26:** 1870–1876.

König, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2012). Protein-RNA interactions: New genomic technologies and perspectives. Nat. Rev. Genet. **13:** 77–83.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene **361:** 13–37.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10:** R25.

Li, F., et al. (2012). Global analysis of RNA secondary structure in two metazoans. Cell Rep. **1:** 69–82.

Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V., and Ding, Y. (2007). Potent effect of target structure on microRNA function. Nat. Struct. Mol. Biol. **14:** 287–294.

Margulies, E.H., Blanchette, M., Haussler, D., and Green, E.D. NISC Comparative Sequencing Program (2003). Identification and characterization of multi-species conserved sequences. Genome Res. **13:** 2507–2518.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17:** 10–12.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature **431:** 343–349.

Montange, R.K., and Batey, R.T. (2008). Riboswitches: Emerging themes in RNA structure and function. Annu. Rev. Biophys. **37:** 117–133.

Muiño, J.M., Kaufmann, K., van Ham, R.C., Angenent, G.C., and Krajewski, P. (2011). ChIP-seq Analysis in R (CSAR): An R package for the statistical detection of protein-bound genomic regions. Plant Methods **7:** 11.

Mustroph, A., Juntawong, P., and Bailey-Serres, J. (2009). Isolation of plant polysomal mRNA by differential centrifugation and ribosome immunopurification methods. Methods Mol. Biol. **553:** 109–126.

Raker, V.A., Mironov, A.A., Gelfand, M.S., and Pervouchine, D.D. (2009). Modulation of alternative splicing by long-range RNA structures in Drosophila. Nucleic Acids Res. **37:** 4533–4544.

Roudier, F., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. EMBO J. **30:** 1928–1938.

Sharp, P.A. (2009). The centrality of RNA. Cell **136:** 577–580.

Tomari, Y., and Zamore, P.D. (2005). Perspective: Machines for RNAi. Genes Dev. **19:** 517–529.

Trappl, K., and Polacek, N. (2011). The ribosome: A molecular machine powered by RNA. Met. Ions Life Sci. **9:** 253–275.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell **147:** 1537–1550.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell **136:** 669–687.

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., and Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. Nat. Rev. Genet. **12:** 641–655.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. **35:** 169–178.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:** 711–716.

Willmann, M.R., Endres, M.W., Cook, R.T., and Gregory, B.D. (2011). The functions of RNA-dependent RNA polymerases in *Arabidopsis*. The Arabidopsis Book **9:** e0146, doi/10.1199/tab.0146.

Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., and Hovorun, D.M. (2003). Downstream elements of mammalian pre-mRNA polyadenylation signals: Primary, secondary and higher-order structures. Nucleic Acids Res. **31:** 1375–1386.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.S., and Gregory, B.D. (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. PLoS Genet. **6:** e1001141.