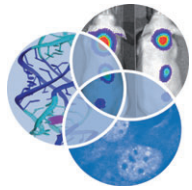
Advanced Review



The potential of the riboSNitch in personalized medicine

Amanda C. Solem,[1] Matthew Halvorsen,[1,2] Silvia B. V. Ramos[3] and Alain Laederach[1,4]*

RNA conformation plays a significant role in stability, ligand binding, transcription, and translation. Single nucleotide variants (SNVs) have the potential to disrupt specific structural elements because RNA folds in a sequence-specific manner. A riboSNitch is an element of RNA structure with a specific function that is disrupted by an SNV or a single nucleotide polymorphism (SNP; or polymorphism; SNVs occur with low frequency in the population, <1%). The riboSNitch is analogous to a riboswitch, where binding of a small molecule rather than mutation alters the structure of the RNA to control gene regulation. RiboSNitches are particularly relevant to interpreting the results of genome-wide association studies (GWAS). Often GWAS identify SNPs associated with a phenotype mapping to noncoding regions of the genome. Because a majority of the human genome is transcribed, significant subsets of GWAS SNPs are putative riboSNitches. The extent to which the transcriptome is tolerant of SNP-induced structure change is still poorly understood. Recent advances in ultra high-throughput structure probing begin to reveal the structural complexities of mutation-induced structure change. This review summarizes our current understanding of SNV and SNP-induced structure change in the human transcriptome and discusses the importance of riboSNitch discovery in interpreting GWAS results and massive sequencing projects.

How to cite this article:
WIREs RNA 2015, 6:517–532. doi: 10.1002/wrna.1291

INTRODUCTION

RNA is now known to be involved in many aspects of genetic regulation. An RNA's function in a cell is determined by not only its primary sequence but also its structure.[1] Unlike DNA, RNA rarely has a complementary second strand, so RNA nucleotides are free to interact in an intramolecular fashion resulting in folding of the polymer chain.[1] Stretches of RNA that are complementary in sequence have a propensity to pair, forming elements of RNA secondary structure.[2] The disruption of functional RNA structural content can be as deleterious to biological function as the disruption of functional sequence content.[1] The consequences of these structural elements depend on their cellular context and may affect alternative splicing,[3] polyadenylation,[4] RNA decay,[5] RNA interference (RNAi),[6,7] regulation by long noncoding RNAs (lncRNAs)[8] or translational efficiency.[9]

The riboSNitch was originally defined as a regulatory element in a messenger, usually located at an untranslated region (UTR) of mRNA or in

*Correspondence to: alain@unc.edu

[1]Department of Biology, University of North Carolina, Chapel Hill, NC, USA

[2]Institute for Genomic Medicine, Columbia University, New York, NY, USA

[3]Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC, USA

[4]Bioinformatics and Computational Biology Program, University of North Carolina, Chapel Hill, NC, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

noncoding RNA (ncRNA), where a single nucleotide polymorphism (SNP) or a single nucleotide variant (SNV) results in an important structural rearrangement.[10–12] It is functionally equivalent to a bacterial riboswitch, an RNA regulatory element in which binding of a small molecule results in a conformational rearrangement that alters gene expression.[13] Riboswitches generally use metabolite binding to switch between two different RNA conformations.[13] Although RiboSNitches could also switch between two RNA conformations upon a single nucleotide change, most RNAs do not homogenously fold to a single structure. Instead, RNA molecules tend to adopt multiple structures.[14,15] Therefore, an SNV or a SNP in a riboSNitch alters the folding landscape, favoring one or multiple alternative conformations.[10,11,16] In this review, we distinguish between SNVs and SNPs; SNPs are common SNVs that occur in >1% of the population.[17] In genetic studies, it is becoming increasingly clear that SNPs alone account for a fraction of phenotypic variation, suggesting that private variants (SNVs) are an important component of disease heritability.[18] That being said, both SNPs and SNVs can and do cause changes to the folding landscape of an RNA, depending on the molecular context. The structural changes to the ensemble in a riboSNitch may be local, disrupting a structural element very close to the SNV (Figure 1(a) and (b)).[1,10,11] Alternatively, an SNV may disrupt distant, long-range interactions that are close in space to the SNV but far away in the sequence. Work with survival motor neuron 2 (SMN2), for example, has shown that intronic long-range interactions close to the splice site and protein-binding sites (Figure 1(e)) have important effects on alternative splicing that contribute to the disease spinal muscular atrophy (SMA).[19] Finally, although a riboSNitch has the potential to alter the global fold of an RNA, experimentally validated riboSNitches to date have had local effects on RNA structure.

Genome-wide association studies (GWAS) aim to determine the relationships between genetic variation and phenotypes.[20] A surprising outcome of these studies is that a vast majority of SNPs associated with human disease phenotypes map outside protein-coding regions of the genome.[11,20,21] Given that a majority of the human genome is transcribed, a majority of SNPs associated with human disease will be transcribed into RNA. Because a subset of these SNPs may affect the structure of these transcripts, there is an important relationship between RNA structure and interpreting the results of GWAS studies. Furthermore, common SNPs can explain only a fraction of human phenotypic variation, suggesting that SNVs play a disproportionate role in determining disease risk in individuals.[18] As a result, accurately predicting and understanding where structural riboSNitches exist and where these may have a phenotypic consequence will likely become an important component of personalized genetic medicine.

RiboSNitches were first proposed following the analysis of the structural consequences of human disease-associated SNVs on UTRs and ncRNAs.[10] RiboSNitches are distinct from many classic and important examples of RNA-based mechanisms of disease such as SNVs that disrupt or alter splicing and expanded repeats.[22] Notably, a riboSNitch must both change RNA structure and involve only a single nucleotide change. Although there are many potential riboSNitches, only a small fraction of SNVs actually significantly change RNA structure. Secondary structure prediction algorithms that return one or several of the most thermodynamically stable (minimum free energy, MFE) structures are not ideal for predicting SNV-induced changes to the structural ensemble. Several algorithms have been developed to predict the impact of SNPs on structural ensembles including SNPfold, RNAsnp, and remuRNA.[10,23,24] In this approach, the full Boltzmann weighted suboptimal structural ensemble is computed for each sequence variant and the base-pairing probabilities per nucleotide calculated. This enables an estimate of structure change between wild type and mutant sequence. Early work with the SNPfold algorithm identified six human diseases including hyperferritinemia cataract syndrome (HCS), β-thalassemia, cartilage-hair hypoplasia (CHH), retinoblastoma, chronic obstructive pulmonary disease (COPD), and hypertension in which more than one associated SNV alters the structure of a UTR or an ncRNA.[10] Three of these riboSNitches were experimentally validated *in vitro*,[11,16,25] suggesting that mutation-induced RNA structure change is an important component of human genetic disease. Several SNVs in the minor spliceosome machinery that cause microcephalic osteodysplastic primordial dwarfism type I (MOPD I) have been extensively validated and act as riboSNitches that interfere with protein binding by changing RNA structure.[26] In addition, recent genome-wide structure probing experiments of the transcriptome allow direct detection of putative structural riboSNitches,[27] as well as the identification of important structural features.[27–31] Together, these results suggest that understanding, measuring, and accurately predicting riboSNitch regulation will be a key to personalized interpretation of genomic sequencing and this review reports on state of the art of riboSNitch detection and validation.

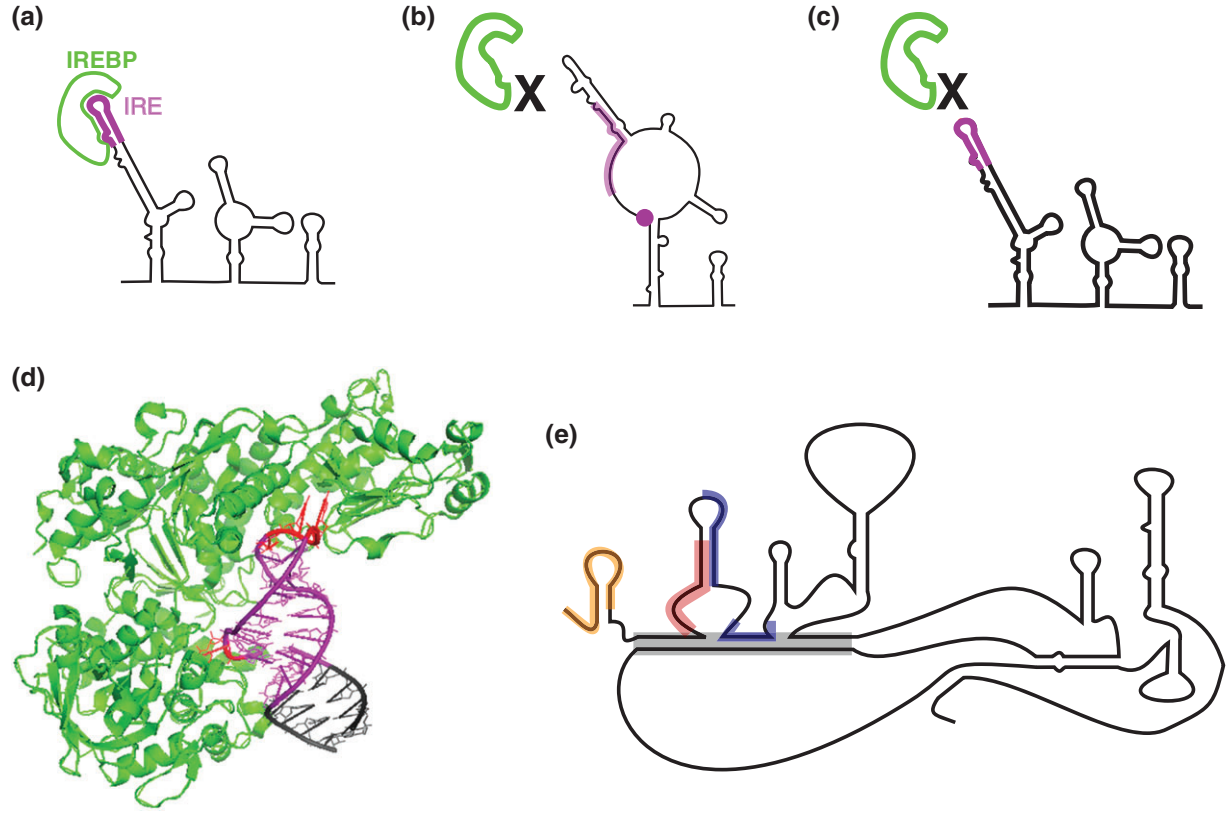


FIGURE 1 | Single nucleotide polymorphisms (SNPs) can affect protein binding through sequence, local, or distant structural changes. (a) The IRE-binding protein (IREBP)/iron responsive element (IRE) interaction in ferritin light chain (FTL) requires both sequence and structural elements. Schematic of productive interaction between IREBP (green) and the IRE (purple) in the human 5′ UTR of ferritin light chain (FTL 5′ UTR) (secondary structure representation in black). (b) The IREBP (green)/IRE (purple) interaction in FTL can be disrupted by mutating a residue (purple dot) that does not contact the protein, but significantly changes the local structure of the IRE. Mutation of any residue that shifts the structure of the IRE could affect binding through this mechanism, including residues that are not actually in the IRE itself. These types of effects are more difficult to predict because they involve accurate secondary structure prediction as well as simple sequence analysis. Secondary structure schematics of the FTL 5′ UTR are based on prediction and experimental data.[11] (c) The IREBP (green)/IRE (purple) interaction is weakened by mutating a nucleotide (red dot) that makes sequence-specific interactions with the protein in the loop region of the IRE even if the mutation does not change the overall structure of the IRE. Mutation of any of the loop and bulge positions that make sequence-specific contacts (red in D) will affect binding through this mechanism. (d) Crystal structure of the rabbit IREBP in complex with frog ferritin H IRE-RNA (3SNP, Ref 33). The IREBP protein (green) has base-specific contacts with the IRE residues in the loop and bulge (red), but nucleotides that create the helical structure (purple) of the IRE also contribute to recognition. This figure was created using PyMOL.[72] (e) Schematic of the secondary structure of the end of exon 7 (orange) and the intron in survival motor neuron 2 (SMN2). Protein-binding sites are shown in red (hnRNP A1/A2B1) and blue (TIA1). Protein binding and exon inclusion are significantly impacted by the long-range interaction shown in gray. This figure was reproduced with modifications.[19] This suggests that riboSNitches might also impact function by changing long-range structural elements.

THE FTL 5′ UTR: A CANONICAL RIBOSNITCH

The riboSNitch located at the 5′ UTR of ferritin light chain (FTL 5′ UTR) mRNA is a classic example of how mutation-induced RNA structure change can result in human disease. FTL codes for a component of the protein Ferritin, which encapsulates excess iron. To regulate gene expression in response to iron, an iron responsive element (IRE) is present in the 5′ UTR of the FTL mRNA; this element is recognized by the IRE-binding protein (IREBP), which inhibits translation of FTL.[32] When iron levels are high, IREBP undergoes a conformational change and cannot bind the IRE, increasing expression of FTL.[32] Mutations that cause HCS disrupt this regulation and cause harmful overexpression of FTL.[32]

The IRE and IREBP interact through a series of highly specific molecular contacts as can be seen in Figure 1(d). The cocrystal structure of the IRE and IREBP clearly reveals the molecular components of the recognition specificity.[33] Highly conserved residues in the hairpin loop and the 5′ bulge

have sequence-specific, direct interactions with the protein. However, the IREBP recognizes not only specific sequences in the RNA, but also the overall fold of the hairpin (Figure 1(a) and (d)). Mutations can disrupt this regulatory interaction in two ways: by directly eliminating sequence-specific contacts (Figure 1(c)) or changing the local structure of the RNA element (Figure 1(b)). We focus here specifically on the case illustrated in Figure 1(b), which cannot be predicted by identification of mutations that abrogate sequence-specific contacts. Instead, the sequence and structure of the RNA element and its surrounding sequences must be considered.

To fully comprehend the FTL riboSNitch, it is important to further describe the concept of suboptimal structural ensembles.[34,35] Although traditional RNA structure prediction algorithms generate a single MFE structure,[2] most RNAs fold into many different conformations that are nearly as energetically stable as the MFE structure. Thermodynamically, this suggests that RNAs will adopt multiple conformations, and that small changes in the folding energy landscape can significantly alter the most likely conformation, which is rarely the MFE.[35,36] Although it is simple to visualize a single structure of the human FTL 5′ UTR, it remains a challenge to represent a structural ensemble. The biggest challenge in visualizing an RNA structural ensemble is that it is fundamentally multidimensional. Principal component analysis (PCA) is often used to reduce multiple dimensional space to two dimensions (as is shown in Figure 2). To generate an interpretable dimensionality reduction, it is imperative to find a two-dimensional PCA space that captures the structurally relevant changes in the ensemble. Most RNAs will adopt multiple conformations, however, the relative number of suboptimal in each cluster of conformation changes. To obtain a useful visualization of the structural ensemble it is essential to provide a PCA algorithm with a well-balanced set of structural clusters, so that the algorithm does not automatically favor structural variation in any one cluster. This is not a trivial computational problem to tackle as it requires a priori knowledge of the space. In ongoing empirical studies, we have found that identifying the mutations that most significantly alter the conformation and pooling their ensembles generally gives an adequate visualization. Nonetheless, ongoing empirical studies in our lab have proven that there is no 'one size fits all' approach to generating these visualizations. In much the same way that creating a visually appealing layout of RNA structure remains a challenge in the field, determining the best PCA decomposition for structural visualization is a contemporary challenge. A Boltzmann suboptimal sampling of RNA structures for the human FTL 5′ UTR is shown in Figure 2.[11] Clustering is performed using PCA and projection onto the first two principal components of the ensemble structure space.[10,11]

This visualization is particularly powerful as it reveals a model for how an SNV can disrupt structure. For the wild-type RNA, a majority of the conformations cluster near the red conformation (Figure 2(a)). An analysis of a representative structure from the red cluster reveals that this conformation folds the IRE into a hairpin (Figure 2(a)). In contrast, the U22G and A56U mutations, identified in two patients with the hyperferritinemia phenotype, alter the ensemble partitioning (Figure 2(b) and (c), respectively). The blue and green conformations become more populated, and in both these cases the IRE does not adopt a hairpin structure and therefore cannot bind the IREBP.

SHAPE Structure Mapping Validates the Predicted Structural Changes in the FTL riboSNitch *in Vitro*

Selective 2′-hydroxyl acylation analyzed by primer extension, or SHAPE[37] chemical probing is one of several techniques that can be used for riboSNitch validation and structural characterization. SHAPE chemistry targets the 2′ hydroxyl of RNA, selectively reacting with flexible nucleotides.[37] Because flexible nucleotides are generally not paired, SHAPE reactivity is inversely correlated with base-pairing probability. Thus, there is a direct correlation between changes in base-pairing probability and the experimentally determined SHAPE reactivity. SHAPE chemical probing can therefore experimentally assess hypotheses about the structural mechanism for disease-associated riboSNitches. In SHAPE experiments, riboSNitches should alter the degree of SHAPE reactivity in areas where they change the structure. In this case, SHAPE data confirm the hypothesis that the U22G SNV significantly alters IRE structure in the FTL 5′ UTR *in vitro* (Figure 3(a)). In contrast, the control mutation (G4A), a common SNP in the FTL 5′ UTR that is not predicted to change structure, is similar to the wild-type RNA with a canonical IRE (Figure 3(b)). Developing and validating riboSNitch hypotheses is therefore dependent on chemical and enzymatic structure probing techniques; the advent of ultra high-throughput techniques for transcriptome-wide probing of RNA structure[27–31,38–44] has begun to revolutionize our understanding of the critical role of RNA structure in controlling function.

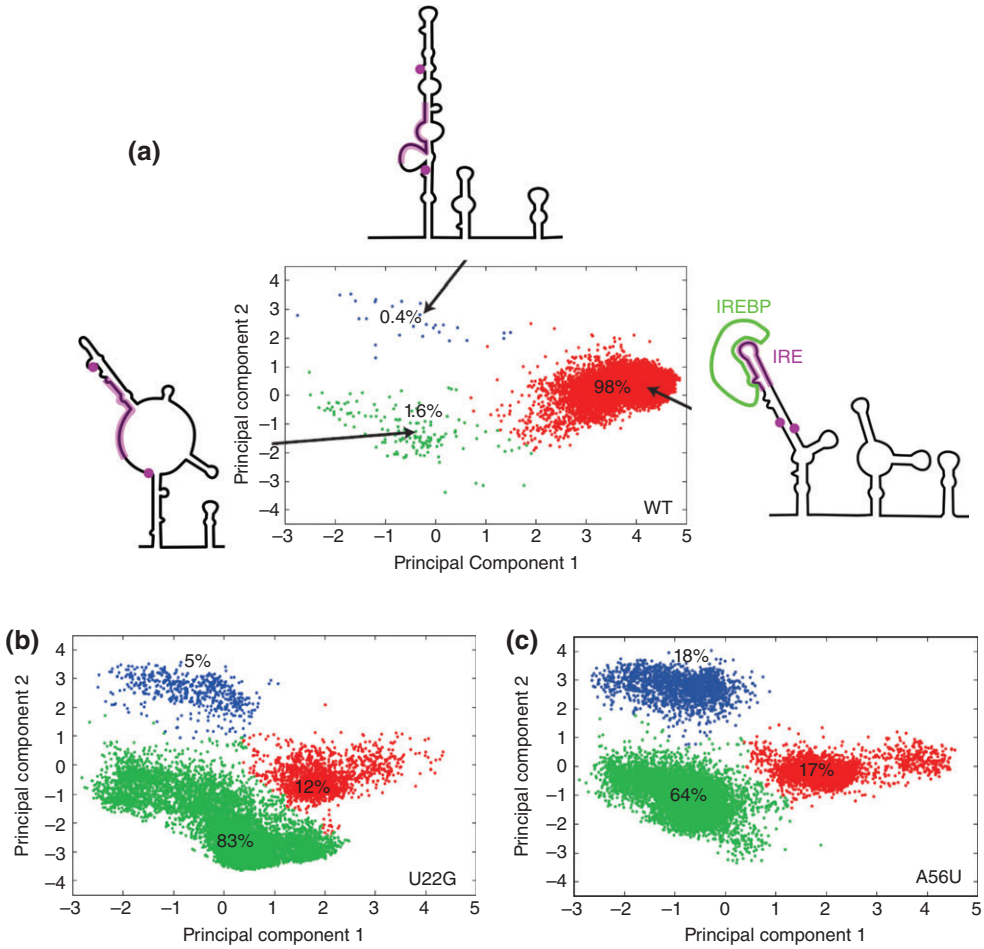


FIGURE 2 | Visualization of how Single nucleotide variants (SNVs) in the ferritin light chain (FTL) riboSNitch are predicted to shift the RNA structural ensemble toward disrupted iron responsive element (IREs). RNA structures were sampled from the Boltzmann suboptimal ensemble predicted for the human FTL 5′ UTR and clustered using principal component analysis (PCA). The 5′ UTR of ferritin light chain (FTL 5′ UTR) forms three distinct clusters of similar structures (red, blue, green). Structures in the red cluster form canonical IREs whereas those in the blue and green clusters do not. The fraction of the population in each cluster is indicated in black. (a) The WT FTL IRE structural ensemble is dominated by the red cluster, indicating that most of the population contains correctly folded IREs. A representative secondary structure schematic for each cluster is shown (black) with the IRE sequence (purple). The position of SNVs associated with hyperferritinemia cataract syndrome residues is indicated on each schematic (purple dots). IRE-binding protein (IREBP) (green) would bind a properly folded IRE (purple). (b) The U22G SNV shifts the structural ensemble away from the structures that can bind IREBP (red cluster) to structures with misfolded IREs and is dominated by structures in the green cluster. (c) The A56U SNV also shifts the structural ensemble toward misfolded IREs that cannot bind IREBP. (Reprinted with permission from Ref 11. Copyright 2012 Cold Spring Harbor Laboratory Press for the RNA Society)

CHH: DELETERIOUS RNAI FROM OFF-TARGET PROCESSING OF MISFOLDED RNA

The association between CHH and the RNA component of mitochondrial RNA processing endoribonuclease (RMRP) is well established—hundreds of variants cause the dwarfism phenotype.[16] However, a mechanistic understanding of the disease etiology is lacking because the precise folding mechanism of the RNA component of RMRP is not well characterized. In particular, the role of an evolutionarily conserved pseudoknot remains puzzling.[16] Several SNVs in RMRP associated with CHH are predicted to be riboSNitches,[10] but without a clear mechanism of disease, it remains unclear how a riboSNitch would affect function.

RNA-seq experiments on liver tissue revealed a surprising pattern of RNA degradation for the lncRNA RMRP. In particular two regions, indicated S1 and S2, were far more abundant than the rest of the RNA (Figure 4(a)), suggesting that these small RNAs are preferentially processed and stabilized. Additional experiments established that RMRP is indeed

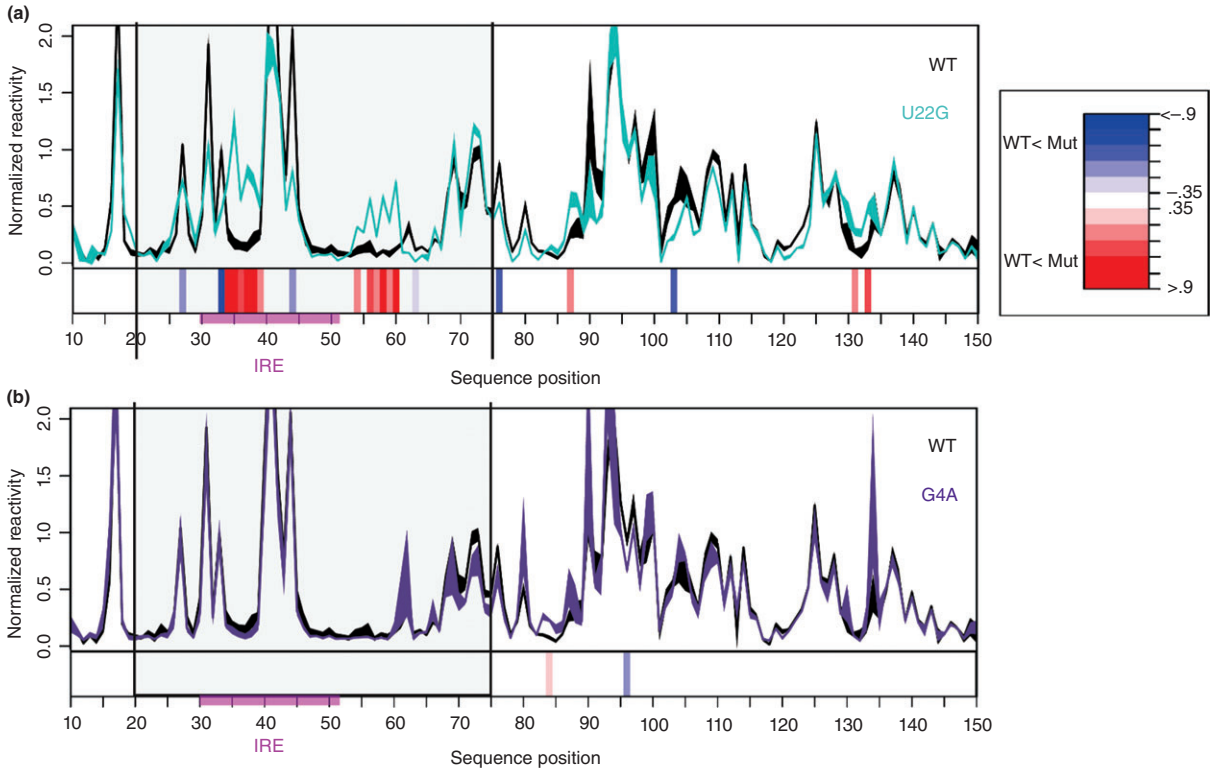


FIGURE 3 | Experimental validation of the ferritin light chain (FTL) riboSNitch. SHAPE experiments were used to probe the secondary structure of the 5′ UTR of ferritin light chain (FTL 5′ UTR). Where the normalized SHAPE reactivity is high, the residue is not base paired. Where the normalized SHAPE reactivity is low, the residue is paired. Differences in SHAPE reactivity between a wild type and mutant version of the RNA are indicated with a heatmap as indicated: blue indicates positions that are more highly modified (single stranded) in the wild type, and red indicates positions that are more highly modified (single stranded) in the mutant. (a) The wild type (WT, black) and mutant (U22G, cyan) forms of the FTL 5′ UTR show significant differences in secondary structure, as predicted, including the IRE (light purple). U22G is associated with hyperferritinemia cataract syndrome (HCS) and the data indicate that the U22G riboSNitch causes disease by changing the structure of the 5′ UTR and disrupting IRE-binding protein (IREBP)/iron responsive element (IRE) interactions. (b) The wild type (WT, black) and mutant (G4A, purple) forms of the FTL 5′ UTR have almost identical SHAPE reactivity, suggesting that the RNA structure does not change. G4A is a single nucleotide polymorphism (SNP) that is neither associated with disease nor predicted to change the RNA structure and serves as a negative control for riboSNitch detection. (Reprinted with permission from Ref 11. Copyright 2012 Cold Spring Harbor Laboratory Press for the RNA Society)

processed by Dicer, a component of the microRNA processing machinery[6] and that overexpressing S1 and/or S2 down regulates genes involved in connective tissue development and skeletal disorders,[16] consistent with them acting as deleterious microRNAs (miRNAs). Importantly, SHAPE experiments indicate that the P2–P4 pseudoknot (yellow structure, Figure 4(b)) is not favored *in vitro*, even though there is strong computational covariance evidence supporting it. Instead, RMRP appears to adopt an alternative structure *in vitro* (Figure 4(a) and blue arcs, Figure 4(b)), where both S1 and S2 appear as paired regions, ideal targets for Dicer. Thus, the SHAPE data suggest that when RMRP misfolds by forming an alternative structure instead of the pseudoknot due to conditions *in vitro* or mutations *in vivo*, it can be deleteriously processed by Dicer.[16] The efficiency of processing RNAs into miRNAs is highly dependent on RNA structure[6] and, as illustrated by RMRP, is another potential source of important riboSNitches, especially in highly abundant lncRNAs like RMRP.

RNA-BINDING PROTEINS AND THE RIBOSNITCH

A vast majority of transcribed RNAs including messenger RNAs (mRNAs) interact with a host of endogenous and exogenous molecules, and in some cases may even assemble into stable complex structures such as ribonucleoproteins (RNPs). Understanding the molecular determinants of RNA/protein interaction specificity and sensitivity is therefore central to a mechanistic understanding of regulatory processes in the cell and accurately predicting the effect of RNA mutations.[1] Many riboSNitches may act by altering

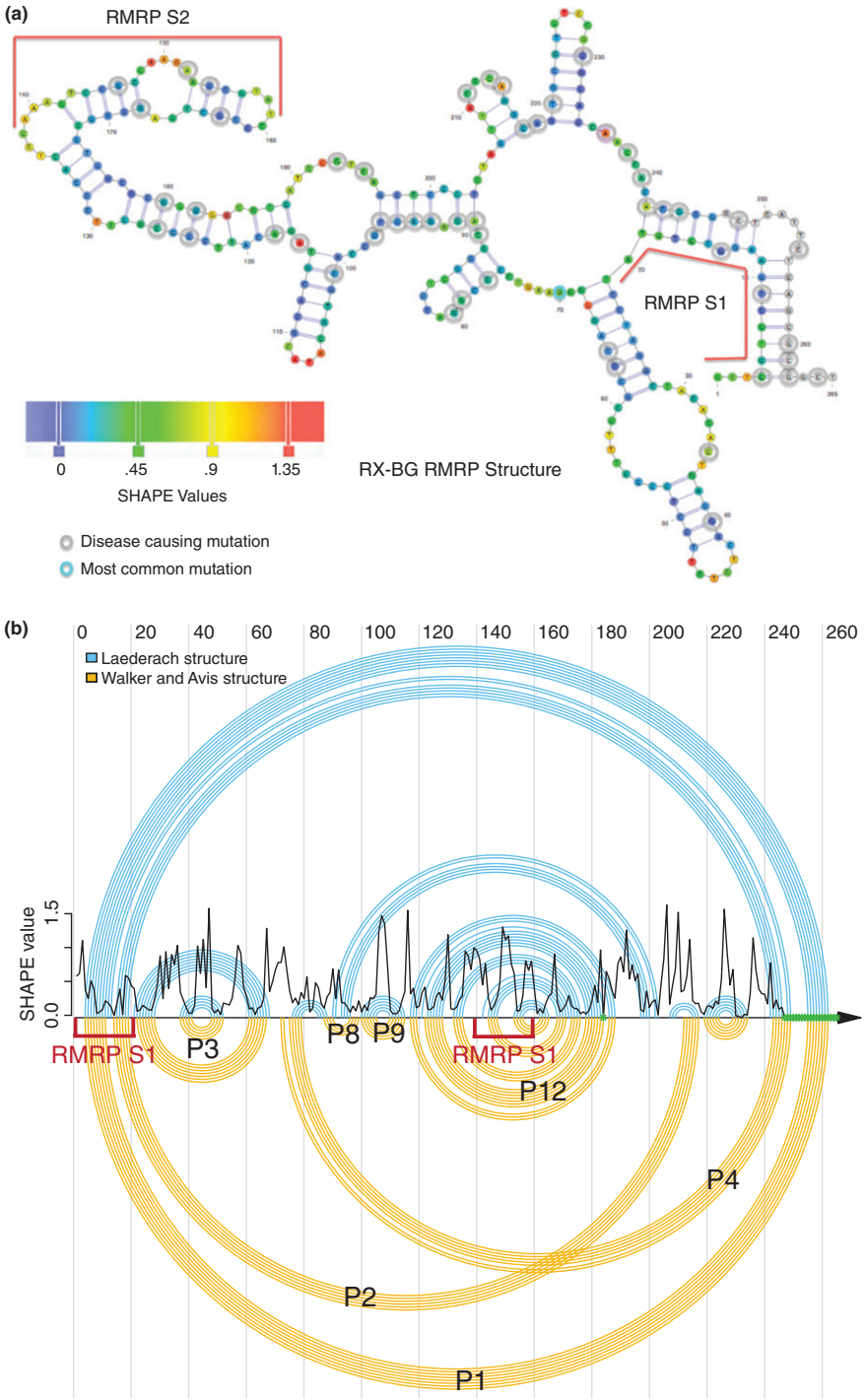


FIGURE 4 | RNA component of mitochondrial RNA processing endoribonuclease (RMRP) forms different structures that are differentially processed into deleterious microRNAs (miRNAs). (a) Secondary structure of RMRP compatible with the experimental SHAPE data. SHAPE reactivity is indicated by color as shown in the heat map: areas of high reactivity (single stranded) are red; areas of low reactivity (base paired) are blue. RMRP-S1 and S2 (red bars) and nucleotides where SNVs are found in CHH patients are indicated (gray circles). (b) Predicted structures based on SHAPE data (blue) and previously published conservation and probing data[73] (yellow) are quite different, indicating heterogeneity in RMRP structure. The R-CHIE R[74] program was used to represent base pairs as arcs. Pseudoknots occur wherever arcs cross; in this case, in the yellow, but not the blue arc diagram. The other major difference is the formation of the alternative P2 helix (Alt-P2) in the blue structure, which resolves the pseudoknot. Taken together, the two predictions are consistent with the hypothesis that RMRP adopts at least two alternative conformations, but it is the experimentally observed blue structure that is more susceptible to Dicer cleavage and processing. (Reprinted with permission from Ref 16. Copyright 2014 Oxford University Press)

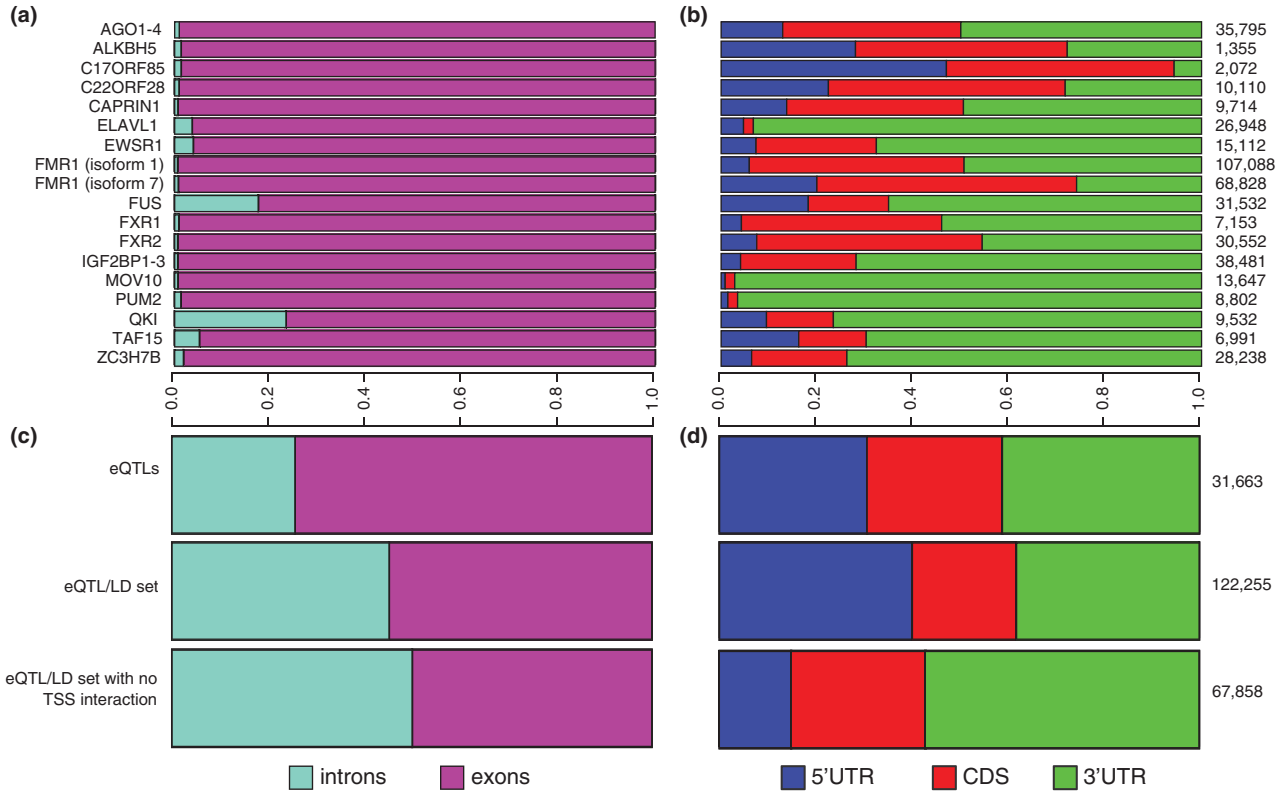


FIGURE 5 | Analysis of photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) and expression Quantitative Trait Loci (eQTLs) data indicate a bias toward RNA–protein binding interactions in 3′ untranslated regions (UTRs). (a and c) PAR-CLIP(a) and eQTL (c) sites are rarely intronic (blue), but are usually exonic (purple). All proteins with PAR-CLIP data at the time of analysis were studied; the name of the protein is indicated next to the bar. Proteins with known roles in splicing (ELAVL1, QK1, and FUS[49–51]) have more intronic sites. (b and d) PAR-CLIP (b) and eQTL (d) sites are also biased to the 3′ UTR (green) as opposed to the 5′ UTR (blue) or coding sequence (CDS) (red). For eQTLs, these effects are more prominent when linkage disequilibrium is taken into account and positions that could directly affect the transcription start site are removed.

RNA/protein binding interactions. As is the case for the FTL IRE/IREBP binding interactions, RNA/RBP interactions are often complex, and the set of sequence and structural features that characterize a binding site vary across the spectrum of RBPs[45,46]. The structure of RNA targets mediates a significant subset of interactions, and disruption of this structural content will have a deleterious effect on binding affinity.[45] Therefore, the molecular context of a riboSNitch is essential to understanding and predicting its function.

To characterize transcriptome-wide binding interactions of RBPs and mRNA targets, we present an overview of recent high-throughput RNA/RBP interaction data derived from photo activatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP). We extracted genomic RNA/RBP interaction sites from 18 different PAR-CLIP datasets available from DORINA, with each dataset representing an RBP or group of homologous RBPs.[47] Here, we present the length-normalized binding density from these data in exons, introns, 5′ and 3′ UTRs, and coding sequence (CDS) of mRNAs (Figure 5). Here, we define exons as sequence in the mature mRNA that can be found in UTRs as well as the CDS.[48] A small subset of proteins (including FUS, ELAVL1, and QK1) significantly binds introns as can be seen in Figure 5(a). This is not surprising because these three proteins have all been shown to play a role in RNA processing and splicing.[49–51] However, the majority of RBPs target exonic regions, including UTRs. Although binding in exonic regions is not homogenous, most RBPs preferentially target 3′ UTRs (Figure 5(b)).

SNVs that are correlated with differential gene expression with an increase in variant dosage are known as expression Quantitative Trait Loci (eQTLs).[52–54] Given that these SNVs are associated with differential RNA expression, it is logical to propose that a subset of eQTL SNVs could be riboSNitches. Despite the appeal of using eQTLs to identify putative functional riboSNitches, a given eQTL SNV may be in linkage disequilibrium (LD) with many others. When SNVs are in LD they are genetically linked and not randomly associated. Thus,

correlation does not imply a single SNV, but an entire haplotype.[11,55] Identifying the causative SNV using eQTLs remains an important challenge in the field.[52–54] It is therefore necessary to consider LD when identifying eQTL SNPs proximal to or directly overlapping with RBP-binding sites on the message. The binding density distributions for eQTLs and eQTLs with passenger LD SNVs derived from the CEU population (Northern Europeans from Utah) of the 1000 genomes project show trends similar to the PAR-CLIP data: most density is in exonic regions, particularly 3′ UTRs (Figure 5). This is consistent with the 3′ UTR preference for a wide range of different RBPs. As expected, when eQTL/passenger SNV LD sets with a variant proximal to the transcription start site (TSS) are excluded, the fraction of eQTLs mapping to the 5′ UTR decreases (Figure 5(d)).

This analysis suggests that there are potential genomic 'hotspots' for riboSNitches, and that these likely occur in 3′ UTRs. eQTLs, by definition, reveal a measurable association between the SNP and gene expression, but most are not directly associated with disease or phenotype. Regulatory networks are known to be robust to some variation, and significant redundancy exists in the cell. Nonetheless, it is well established that genetic predisposition exists, diseases do 'run' in families. Interestingly, it is often impossible to identify a single SNP that is highly predictive of a human trait. Thus, although most riboSNitches individually may not directly cause disease, the wrong combination of riboSNitch-induced deregulation will likely play an important role in many complex traits. The ability to identify riboSNitches is thus a major component of interpretation of personalized genomic sequencing.

Characterized mRNA Elements Likely to Harbor RiboSNitches

Many functional riboSNitches in mRNAs are likely to exert their effect by lowering the accessibility to RBPs or regulatory RNAs due to changes in the binding site of a target RNA. Recent *in silico* analyses of RNA-binding preferences for a subset of RBPs clearly indicate that structure plays a central role in determining the binding affinity of some RBPs.[56] A majority of RBPs with well-characterized binding preferences have a high preference for single-stranded RNA, and are observed to have a significantly lower binding affinity for structured RNA, even if the region contains the sequence motif critical for binding.[45,57] Accessibility is also a key variable in controlling mRNA/miRNA binding interactions.[58] Thus, a functional riboSNitch can shift the underlying structural ensemble toward a state of greater stability and less entropy, preventing binding.

Alternatively, as is the case for the FTL riboSNitch, disrupting structural elements critical to recruitment of particular RBPs is causative. For a subset of RBPs, folded structure is a critical determinant of binding affinity.[1] In addition, while the RBPs analyzed in Figure 5 prefer single-stranded structure, there are also a host of interesting examples of structure-dependent RNA/RBP interactions. For instance, SBP2 binds the Selenocysteine Insertion sequence element (SECIS), an RNA structure that facilitates the incorporation of a Selenocysteine amino acid instead of termination during translation.[59] Disruption of this structure will trigger premature translation termination.[60] The mRNA-editing protein ADAR also binds stable stem regions of RNA structure[61]; thus, a riboSNitch can potentially reduce RNA editing. In summary, the nature of the structural change is central to predicting the type of interaction that can be affected by a riboSNitch.

TRANSCRIPTOME-WIDE STRUCTURE PROBING TO DISCOVER PUTATIVE RIBOSNITCHES

The first riboSNitches were predicted using a forward genetics approach: causative private SNVs associated with single gene heritable genetic diseases were evaluated for their ability to disrupt RNA structure.[10,11] Given that these SNVs cause the disease in these individuals, experimentally validating the structure change with traditional chemical and enzymatic structure probing established them as functional riboSNitches.[10,16,25] The advent of transcriptome-wide RNA structure studies incorporating chemical and enzymatic structure probing protocols for read out by next-generation sequencing[27–31,39,41–44] is fundamentally changing the way riboSNitches are studied (Box 1). RiboSNitches are now examined using a reverse genetics approach, quickly identifying structural riboSNitches experimentally for SNVs with no known phenotype. In a recent transcriptome-wide analysis of RNA structure using parallel analysis of RNA structure (PARS), the Chang lab identified 1907 such structural riboSNitches.[27] This result suggests that SNVs readily affect transcriptome structure, and that these changes are in most cases benign. It is nonetheless important to remember that the genetics of complex disease predisposition in humans is still poorly understood, and a lack of functional annotation for an SNV does not mean it is not functional.

BOX 1

BROAD-RANGING IMPACTS OF NEXT-GENERATION SEQUENCING FOR RIBOSNITCHES

Next-generation sequencing techniques are significantly changing research methods and shaping the future of personalized medicine. For the case of riboSNitches, next-generation sequencing has been revolutionary in many ways. For example, the increase in human genome sequencing has greatly expanded our understanding of human genetic variation and the subset that contributes to disease. These techniques are also significantly changing the types of questions that can be asked about RNA structure. Previously, it was a significant undertaking to validate a few candidate riboSNitches; now, researchers can compare the structures of entire transcriptomes of multiple organisms or individuals. This paradigm shift is changing how riboSNitches are identified. *Finally, as next-generation sequencing continues to drive down sequencing costs and becomes more prominent in the clinic, it is also changing the future of personalized medicine. Discovering the impact of riboSNitches will be important in interpreting SNVs in personalized medicine.*

As a result, accurately predicting the subset of structural riboSNitches from an individual's genomic sequence is likely to be an important component of predicting predisposition for complex genetic diseases. As a mutation is an element of the riboSNitch, the ability to identify mutations that alter RNA structure on a transcriptome-wide scale is an important aspect of reconciling the large compendium of genetic data now available on human diseases and phenotypes[20,21,62,63] and determining the mutations for which RNA structure change is a component of the phenotypic etiology. Two new techniques, SHAPE-MaP and RING-MaP, detect sites of modification by mutation rather than strand termination thereby eliminating several types of artifacts.[39,43] These two methods should generally not interfere with identification of true natural variants despite the fact that they induce mutation, as they include negative controls and the induced mutation rate is low.

Three of these new transcriptome-wide studies use chemical modifiers such as dimethyl sulfate (DMS) and *N*-cyclohexyl-*N*′-(2-morpholinoethyl)carbodiimide (CMCT) while the others use enzymatic cleavage to infer the RNA structure.[27–31] DMS is a classic tool for probing RNA structure as it strongly modifies solvent accessible and unpaired adenosine (A), cytosine (C), or guanosine (G) residues; however, it cannot react with residues involved in Watson-Crick base pairing.[29–31] In this method, when the reverse transcriptase (RT) reaches a DMS-modified A or C base, it terminates the extension. Traditionally, these sites are identified by gel or capillary electrophoresis.[29–31] As DMS is a small molecule that can diffuse across membranes, it is also very useful for modifying RNA *in vivo* as well as *in vitro*.[29–31] CMCT also modifies RNA nucleotides that are single stranded rather than base paired, but it is specific for uridine (U) and guanosine (G); during reverse transcription, any of these mutations cause termination.[31] Enzymatic cleavage of RNA is also a proven structure probing technique.[27,28] It is based on the differential ability of RNase V1 and S1 nuclease to cleave double- and single-stranded RNA, respectively.[27,28] V1 and S1 are not sequence or nucleotide specific.[27,28] The disadvantage is that these enzymes are large, so they cannot diffuse across the cell membrane and thus have limited use to obtain structural data on RNA *in vivo*.[27,28] Note that when any of these techniques are used to probe structural ensembles, positions that are single stranded in some RNA molecules and double stranded in others are likely to appear to be intermediately paired.

The two studies using primarily DMS structure probing use different model organisms: one used Arabidopsis seedlings[30] while the other studied yeast and mammalian cells.[29] In the former study,[30] the experiments included negative controls without DMS, allowing removal from analysis of degradation products and natural stops detected during the RT reaction. This technique was validated using a known RNA structure, 18S rRNA. The results generated under these conditions demonstrated several global features in RNA structure in Arabidopsis. First, they found that the 5′ UTR of most mRNAs near the start codon have reduced structure, which is consistent with previous *in silico* and *in vitro* findings.[28,64,65] Collectively, this supports the idea that reduced structure could facilitate translation. Second, they found a periodic structural triplet repeat in the CDS that was most predominant in highly translated genes and absent in poorly translated transcripts, suggesting that the structure of the CDS may have a role in translational efficiency, consistent with previous data and predictions.[28,65,66] Finally, they found a pattern of structured followed by unstructured regions near polyadenylation sites suggesting that this not yet fully characterized pattern of RNA structure may be involved in alternative

polyadenylation. In light of these results, there are three possible ways that riboSNitches might alter the fate of an RNA transcript by changing its structure—altering how open the 5′ UTR is near the translation start site to modify translation initiation, disrupting, or promoting periodic structure in the CDS to modulate translation efficiency, or altering structural elements in the 3′ UTR to modify the choice of polyadenylation sites.

In contrast, the later study[29] used a different strategy, modifying the RNA *in vivo* and *in vitro*. *For in vitro* experiments, the extracted RNA was either denatured or refolded before addition of DMS to modify the RNA. This allowed a direct comparison of the *in vivo* and *in vitro* RNA structures as well as estimating the maximum reactivity of each residue. This technique was also validated using a known ribosomal RNA structure as well as several known mRNA structures. Interestingly, the results suggested that most RNAs are less folded *in vivo* than *in vitro*, partially due to the action of ATP-dependent proteins such as helicases. However, in a few cases, the *in vivo* structure appears to be more stably folded. The authors propose that the RNAs in this category may have important structural roles. Indeed, this appears to be the case for three stem-loop structures, two in 5′ UTRs and one in a 3′ UTR, validated using a reporter system. In summary, these studies not only provide methods to produce data that can be used to experimentally to identify riboSNitches but also raise new hypotheses on how RNA structural elements, or their disruption by riboSNitches, may impact gene expression.

Very recently, a new technique, chemical inference of RNA structures followed by massive parallel sequencing (CIRS-Seq) was applied to the transcriptome of mouse embryonic stem cells.[31] CIRS-Seq combines DMS, CMCT, and a negative control in parallel reactions to probe RNA structure. The authors gently extract and deproteinize the RNA from cells to try to separate RNA structure from protein-binding sites. By combining DMS (A/C) and CMCT (G/U) data into one experiment, they are able to probe all four nucleotides. The experiments were validated using known tRNA structures. Surprisingly, unlike Arabadopsis where overall more structure in the coding region was observed relative to the UTRs,[30] the CRIS-seq data suggest that regions near the end of the 5′ UTR and the beginning of the 3′ UTR are actually more structured than the CDS.[31] They do, however, observe more open structure near the start and stop codons, perhaps facilitating binding and release of the ribosome. Note that the Arabidopsis study performed modification *in vivo* in the presence of proteins,[30] whereas this work modified deproteinized, extracted RNA from mammalian cells.[31] In addition, they also observe the periodic triplet repeat in highly translated CDS.[30,31] The data also show that small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), tRNAs, and long intergenic noncoding RNAs (lincRNAs) are generally more structured than mRNAs. Finally, they observe a slightly higher RNA accessibility at the target binding sites of the RNA-binding protein Lin28a, consistent with predictions that Lin28a target sites are single stranded.[31,67]

PARS using V1 and S1 nuclease to map the structure of either refolded or deproteinized RNA from yeast and lymphoblast cell lines provides the most direct evidence for the importance of the riboSNitch.[27,28] In initial work in yeast, RNA was extracted and refolded before being digested in parallel with V1 and S1 nucleases.[28] The technique was validated by including regions of known, exogenous RNAs as a positive control. The authors noted that the CDS is generally more structured than UTRs, start and stop codons are very accessible, with less secondary structure especially near the start codon. They first observed the triplet repeat periodicity structural signature in CDS associated with efficient translation, consistent with predictions and later results from other laboratories.[30,65,66] The data obtained from human cells using this analysis are in accordance to several global features of the RNA structure found in Arabidopsis and the original study in yeast, such as the open structure near the start codons and a periodic three-nucleotide structure in the CDS.[28,30,64–66] They also observed that a subset of sites are consistently different between refolded and native deproteinized samples, suggesting that these are putative candidates for RNA structure regulation *in vivo*. They noted an asymmetric structural signature at exon–exon junctions and showed that bonafide miRNA targets are located upstream of structurally accessible areas, while sequences containing potential miRNA binding sites that are not actually functional lack this accessibility.

Most importantly, this last study compared multiple lymphoblast cell lines including a family trio with a mother, father, and child allowing them to identify and validate a set of putative structural riboSNitches by comparing structural data from individuals with different alleles.[27] They found that 1907 of 12,233 SNVs in the trio were putative riboSNitches, i.e., there is a measurable change in PARS scores near the SNV. Of these putative riboSNitches, 211 are eQTLs and 22 are SNPs associated with human disease and phenotypes through GWAS. SNVs changing structure near observed splice junctions are also

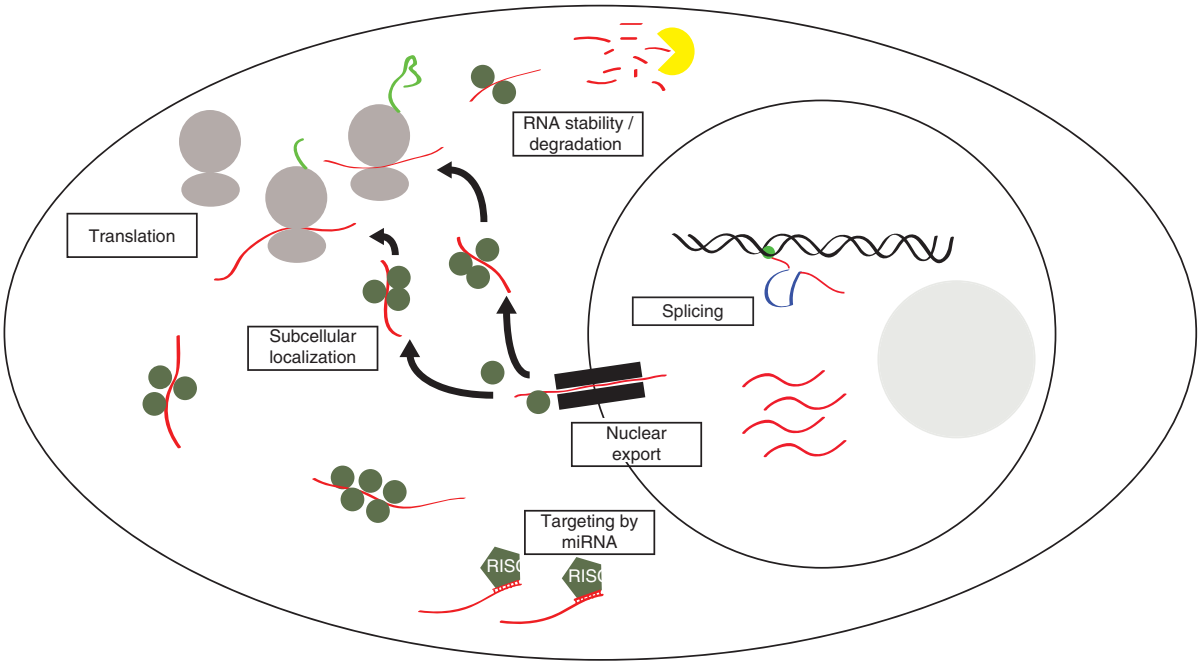


FIGURE 6 | Changes in RNA structure could contribute to regulation of gene expression in a variety of ways. As RNA structure is vital for function, changes in RNA structure could either promote or inhibit gene expression. A conformational change could alter diverse processes including splicing, export, localization, translation initiation or elongation, or decay. Conformational changes may also have a significant role in RNA-based mechanisms of regulation that involve noncoding RNA (ncRNA) such as microRNAs (miRNAs) or long noncoding RNAs (lncRNAs). A given riboSNitch need only impact a single event or interaction during its lifetime to have a real impact on gene expression. Elucidating the impact of riboSNitches on the transcriptome and genetic regulation will be vital to a mechanistic understanding of disease and the future of personalized medicine.

correlated with changes in alternative splicing, suggesting a new molecular mechanism for this subset of putative riboSNitches. In fact, recent analysis of RNA-seq data from 465 of the thousand genomes lines has identified hundreds of SNPs as splicing QTLS (sQTLs), SNPs that clearly alter patterns of alternative splicing.[68] The set of sQTLs provides an additional source of potentially new riboSNitches for further analysis. In addition, the data suggest that covariation of some SNVs may have a role in maintaining functional RNA structure, consistent with previous analysis.[11] SNVs that alter structure were found to be less likely to occur in 3′ UTRs, predicted miRNA target sites, and protein-binding sites. At first, this result might appear to contradict the data presented in Figure 5. However, these PARS experiments were performed using samples from healthy individuals. It is therefore likely that a majority of these SNVs have limited functional consequence, but occur in sites where structure change is tolerated. As such, it is expected and consistent with the data shown in Figure 5 that these putative riboSNitches are less prevalent in 3′ UTRs, as structure change is more likely to affect posttranscriptional regulation in these regions of messages. It has yet to be experimentally determined whether the sensitivity to mutation of RNAs evolved to adopt a single conformation to exercise their function is higher, lower or equal to those which do not need to fold to a single conformation. Computationally, ensembles without structural selection are predicted to be easily disrupted by mutation.[69] It is likely that the nature of the structural disruption, and the specific structures being disrupted (or stabilized) in a riboSNitch will be integral to determining if there is a strong phenotypic effect.

CONCLUSION

The overall impact and role of RNA structure and riboSNitches is not clearly understood in mRNAs, but there are many structural elements in mRNAs that have clear functional roles. Like other types of RNA, mRNA is often targeted for binding by proteins or ncRNAs. The binding affinity of these biomolecules for target mRNA is known to be dependent on the accessibility of the target site and the RNA's folded structure.[58,70] As detailed in Figure 6 there are multiple regulatory events that occur during an mRNA's lifetime that are dependent on these interactions. The number of interactions guiding the posttranscriptional processes illustrated in Figure 6 is significant and highly dependent on molecular

context. Structure is one important and understudied component of epitranscriptomics, the role of RNA modification and posttranscriptional regulation in general. Only an integrative approach can begin to reconcile the genomic data to facilitate the development of predictive models for the processes illustrated in Figure 6.

A significant outcome of more recent GWAS studies is that a majority of SNPs associated with more complex human phenotypes map to noncoding regions of the genome. The associated SNP is not necessarily causative, however. Nonetheless, the riboSNitch as a concept offers an attractive hypothesis for reconciling GWAS data. Accurate prediction of riboSNitches therefore has the potential to facilitate the identification of causative SNPs in large LD blocks. Even more importantly, common SNPs can only explain a subset of observed human phenotypic variation; it is likely the SNVs (or private variants) play a large role in determining genetic predisposition. As such, riboSNitch detection and prediction for private SNVs will likely play an important role in the personalized interpretation of genome-wide sequencing. Although current work has focused on heritable, germline SNVs that are associated with disease, riboSNitches that arise from somatic mutations may also play important roles in disease. Indeed, one of the original predicted riboSNitches arises from germline mutations in the tumor suppressor retinoblastoma 1 (RB1) and is associated with retinoblastoma.[10,25] Similarly, riboSNitches that arise from somatic mutations will likely drive cancer.

Transcriptome-wide probing experiments are revealing a second layer of information in RNA: its structure. The fact that 1907 riboSNitches were detected in a single triplet comparison of human cell lines[27] suggests not all structure in the transcriptome has a specific function. Nonetheless, the fact that SNVs in UTRs and ncRNAs such as FTL and RMRP do cause human disease[11,16] suggests a subset of riboSNitches will shape the human phenotypic landscape. Functional riboSNitches are also potential therapeutic targets and could be targeted, for example, by small molecules specific to the disease-associated riboSNitch[71]. The degree to which structure change will be predictive of function remains to be determined; clearly the context of the riboSNitch will be an important component of its function, as well as the nature of the structural change. In addition, although many riboSNitches may not have measurable effects in isolation, combinations of multiple riboSNitches in individual genetic backgrounds may make important contributions to complex diseases. Therefore, the importance of RNA structure and the riboSNitch will likely increase as medicine moves toward personalization.

ACKNOWLEDGMENTS

This work is supported by U.S. National Institutes of Health grants NHLBI R01HL111527 and NIGMS R01GM101237 to A.L. and NICHD K08HD069597 to S.B.R.

REFERENCES

1. Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* 2004, 5:908–919.
2. Zuker M, Sankoff D. RNA secondary structures and their prediction. *Bull Math Biol* 1984, 46:591–621.
3. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003, 72:291–336.
4. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* 2013, 14:496–506.
5. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* 2012, 13:246–259.
6. Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol* 2007, 8:23–36.
7. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009, 136:215–233.
8. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell* 2013, 154:26–46.
9. Kuersten S, Radek A, Vogel C, Penalva LO. Translation regulation gets its 'omics' moment. *WIREs RNA* 2013, 4:617–630.
10. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet* 2010, 6: e1001074.
11. Martin JS, Halvorsen M, Davis-Neulander L, Ritz J, Gopinath C, Beauregard A, Laederach A. Structural effects of linkage disequilibrium on the transcriptome. *RNA* 2012, 18:77–87.

12. Ritz J, Martin JS, Laederach A. Evaluating our ability to predict the structural disruption of RNA by SNPs. *BMC Genomics* 2012, 13(Suppl 4):S6.
13. Breaker RR. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* 2012, 4:a003566.
14. Ritz J, Martin JS, Laederach A. Evolutionary evidence for alternative structure in RNA sequence co-variation. *PLoS Comput Biol* 2013, 9:e1003152.
15. Solomatin SV, Greenfeld M, Chu S, Herschlag D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 2010, 463:681–684.
16. Rogler LE, Kosmyna B, Moskowitz D, Bebawee R, Rahimzadeh J, Kutchko K, Laederach A, Notarangelo LD, Giliani S, Bouhassira E, et al. Small RNAs derived from lncRNA RNase MRP have gene-silencing activity relevant to human cartilage-hair hypoplasia. *Hum Mol Genet* 2014, 23:368–382.
17. Haraksingh RR, Snyder MP. Impacts of variation in the human genome on gene regulation. *J Mol Biol* 2013, 425:3970–3977.
18. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012, 337:100–104.
19. Singh NN, Lawler MN, Ottesen EW, Upreti D, Kaczynski JR, Singh RN. An intronic structure enabled by a long-distance interaction serves as a novel target for splicing correction in spinal muscular atrophy. *Nucleic Acids Res* 2013, 41:8144–8165.
20. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, Kathiresan S, Keaney JF Jr, Keyes MJ, Lin JP, et al. Genome-wide association with select biomarker traits in the Framingham Heart Study. *BMC Med Genet* 2007, 8(suppl 1):S11.
21. Bulik-Sullivan B, Selitsky S, Sethupathy P. Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies. *Hum Mutat* 2013, 34:1049–1056.
22. Douglas AG, Wood MJ. RNA splicing: disease and therapy. *Brief Funct Genomics* 2011, 10:151–164.
23. Salari R, Kimchi-Sarfaty C, Gottesman MM, Przytycka TM. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic Acids Res* 2013, 41:44–53.
24. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. *Hum Mutat* 2013, 34:546–556.
25. Kutchko K, Sanders W, Ziehr B, Phillips G, Solem A, Halvorsen M, Weeks K, Moorman N, Laederach A. Multiple conformations are a conserved and regulatory feature of the RB1 5′UTR. *RNA* 2015, 21:1–12.
26. Jafarifar F, Dietrich RC, Hiznay JM, Padgett RA. Biochemical defects in minor spliceosome function in the developmental disorder MOPD I. *RNA* 2014, 20:1078–1089.
27. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 2014, 505:706–709.
28. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 2010, 467:103–107.
29. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 2014, 505:701–705.
30. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 2014, 505:696–700.
31. Incarnato D, Neri F, Anselmi F, Oliviero S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol* 2014, 15:491.
32. Rouault TA. The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat Chem Biol* 2006, 2:406–414.
33. Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla-Camps JC, Theil EC, Volz K. Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* 2006, 314:1903–1908.
34. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 2004, 32:W135–W141.
35. Ding Y, Chan CY, Lawrence CE. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* 2005, 11:1157–1166.
36. Freyhult E, Gardner PP, Moulton V. A comparison of RNA folding measures. *BMC Bioinformatics* 2005, 6:241.
37. Wilkinson KA, Merino EJ, Weeks KM. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc* 2005, 127:4659–4667.
38. Ramos SB, Laederach A. Molecular biology: a second layer of information in RNA. *Nature* 2014, 505:621–622.
39. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 2014, 11: 959–965.
40. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. 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.

41. Talkish J, May G, Lin Y, Woolford JL Jr, McManus CJ. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 2014, 20:713–720.
42. Kielpinski LJ, Vinther J. Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res* 2014, 42:e70.
43. Homan PJ, Favorov OV, Lavender CA, Kursun O, Ge X, Busan S, Dokholyan NV, Weeks KM. Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci USA* 2014, 111:13858–13863.
44. Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP, Pachter L. Modeling and automation of sequencing-based characterization of RNA structure. *Proc Natl Acad Sci USA* 2011, 108:11069–11074.
45. Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* 2014, 54:887–900.
46. Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* 2009, 27:667–670.
47. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* 2012, 40:D180–D186.
48. Zhang MQ. Statistical features of human exons and their flanking regions. *Hum Mol Genet* 1998, 7:919–932.
49. Rogelj B, Easton LE, Bogu GK, Stanton LW, Rot G, Curk T, Zupan B, Sugimoto Y, Modic M, Haberman N, et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci Rep* 2012, 2:603.
50. Ryder SP, Williamson JR. Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination. *RNA* 2004, 10:1449–1458.
51. Uren PJ, Burns SC, Ruan J, Singh KK, Smith AD, Penalva LO. Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *J Biol Chem* 2011, 286:37063–37066.
52. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009, 325:1246–1250.
53. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet* 2011, 27:72–79.
54. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* 2012, 8:e1002639.
55. Morton NE. Into the post-HapMap era. *Adv Genet* 2008, 60:727–742.
56. Zhao H, Yang Y, Zhou Y. Prediction of RNA binding proteins comes of age from low resolution to high resolution. *Mol Biosyst* 2013, 9:2417–2425.
57. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 2013, 499: 172–177.
58. 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.
59. Low SC, Grundner-Culemann E, Harney JW, Berry MJ. SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy. *EMBO J* 2000, 19:6882–6890.
60. Latreche L, Jean-Jean O, Driscoll DM, Chavatte L. Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. *Nucleic Acids Res* 2009, 37:5868–5880.
61. Bass BL, Weintraub H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 1988, 55:1089–1098.
62. Lee SH, van der Werf JH, Hayes BJ, Goddard ME, Visscher PM. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet* 2008, 4:e1000231.
63. Nuinoon M, Makarasara W, Mushiroda T, Setianingsih I, Wahidiyat PA, Sripichai O, Kumasaka N, Takahashi A, Svasti S, Munkongdee T, et al. A genome-wide association identified the common genetic variants influence disease severity in $\beta(0)$-thalassemia/hemoglobin E. *Hum Genet* 2009, 127:303–314.
64. Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD. Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *Plant Cell* 2012, 24:4346–4359.
65. 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.
66. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009, 324:218–223.
67. Cho J, Chang H, Kwon SC, Kim B, Kim Y, Choe J, Ha M, Kim YK, Kim VN. LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* 2012, 151:765–777.
68. Monlong J, Calvo M, Ferreira PG, Guigo R. Identification of genetic variants associated with alternative splicing using sQTLseekeR. *Nat Commun* 2014, 5: 4698.

69. Ancel LW, Fontana W. Plasticity, evolvability, and modularity in RNA. *J Exp Zool* 2000, 288:242–283.

70. Kazan H, Ray D, Chan ET, Hughes TR, Morris Q. RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput Biol* 2010, 6:e1000832.

71. Guan L, Disney MD. Small-molecule-mediated cleavage of RNA in living cells. *Angew Chem Int Ed Engl* 2013, 52:1462–1465.

72. The PyMOL Molecular Graphics System, Version 1.3r1, Schrödinger, LLC. 2010.

73. Thiel CT, Rauch A. The molecular basis of the cartilage-hair hypoplasia-anauxetic dysplasia spectrum. *Best Pract Res Clin Endocrinol Metab* 2011, 25:131–142.

74. Lai D, Proctor JR, Zhu JY, Meyer IM. R-CHIE: a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Res* 2012, 40:e95.

