Concordant dysregulation of miR-5p and miR-3p arms of the same precursor microRNA may be a mechanism in inducing cell proliferation and tumorigenesis: a lung cancer study

RAMKRISHNA MITRA,¹ CHEN-CHING LIN,¹ CHRISTINE M. EISCHEN,² SANGHAMITRA BANDYOPADHYAY,³ and ZHONGMING ZHAO^{1,4,5,6}

¹Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA

²Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

³Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

⁴Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

⁵Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee 37212, USA

⁶Center for Quantitative Sciences, Vanderbilt University, Nashville, Tennessee 37232, USA

ABSTRACT

A precursor microRNA (miRNA) has two arms: miR-5p and miR-3p (miR-5p/-3p). Depending on the tissue or cell types, both arms can become functional. However, little is known about their coregulatory mechanisms during the tumorigenic process. Here, by using the large-scale miRNA expression profiles of five cancer types, we revealed that several of miR-5p/-3p arms were concordantly dysregulated in each cancer. To explore possible coregulatory mechanisms of concordantly dysregulated miR-5p/-3p pairs, we developed a robust computational framework and applied it to lung cancer data. The framework deciphers miR-5p/-3p coregulated protein interaction networks critical to lung cancer development. As a novel part in the method, we uniquely applied the second-order partial correlation to minimize false-positive regulations. Using 279 matched miRNA and mRNA expression profiles extracted from tumor and normal lung tissue samples, we identified 17 aberrantly expressed miR-5p/-3p pairs that potentially modulate the gene expression of 35 protein complexes. Functional analyses revealed that these complexes are associated with cancer-related biological processes, suggesting the oncogenic potential of the reported miR-5p/-3p pairs. Specifically, we revealed that the reduced expression of miR-145-5p/-3p pair potentially contributes to elevated expression of genes in the “FOXM1 transcription factor network” pathway, which may consequently lead to uncontrolled cell proliferation. Subsequently, the regulation of miR-145-5p/-3p in the FOXM1 signaling pathway was validated by a cohort of 104 matched miRNA and protein (reverse-phase protein array) expression profiles in lung cancer. In summary, our computational framework provides a novel tool to study miR-5p/-3p coregulatory mechanisms in cancer and other diseases.

Keywords: concordant dysregulation; lung cancer; miR-5p/-3p arms; protein array; protein complex

INTRODUCTION

MicroRNAs (miRNAs) are small (~21–23 nt) noncoding RNAs that are abundant in cells and function as crucial regulators of gene expression. The canonical model of miRNA biogenesis is tightly controlled by multiple enzymes that produce three major RNA products: primary (pri-), precursor (pre-), and mature miRNA (Bartel 2004). Mature miRNA originates from the 5' arm or the 3' arm of the precursor product and is denoted with a -5p or -3p suffix, respectively (Kozomara and Griffiths-Jones 2014). One of the arms is incorporated into the RNA induced silencing complex (RISC) and becomes functional, whereas the other arm is generally

considered a byproduct and is typically degraded (Supplemental Fig. S1; scenario 1). However, recent studies have found that, depending on the tissue or cell-type, both mature miR-5p and miR-3p arms (hereafter referred to as miR-5p/-3p pair) of a pre-miRNA can be associated with the RISC (Supplemental Fig. S1; scenario 2) and implicated in the pathology of several types of cancer through coregulating common set of genes (Chang et al. 2013; Uchino et al. 2013; Yang et al. 2013). For example, miR-582-5p and miR-582-3p (miR-582-5p/-3p) pair targets three mRNA—*PGGT1B*, *LRK2*, and *DIXDC1*, and controls high-grade bladder cancer

Corresponding author: zhongming.zhao@vanderbilt.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.048132.114>.

© 2015 Mitra et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

progression in vitro and in vivo (Uchino et al. 2013); and miR-17-5p/-3p pair targets *TIMP3* and induces prostate tumor growth and invasion (Yang et al. 2013). Thus far, evidence has been obtained only through the study of specific miRNA pairs. Therefore, a robust computational framework using large-scale clinical data is required to advance our understanding of miR-5p/-3p-mediated critical coregulatory mechanisms in the etiology of cancer.

Motivated by the above experimental discoveries, we developed a novel computational framework and applied it to lung cancer to elucidate the critical roles of miR-5p/-3p pairs in its pathological process. This framework deciphers the cooperative regulation ability of miR-5p/-3p pairs under the premises that (1) coexpressed genes are more likely to be coregulated (Ideker et al. 2001; Amar et al. 2013), and (2) the expression correlation between two genes may reflect an induced association due to a mutual association with regulating molecule(s) (Barzel and Barabasi 2013). If the degree of coexpression between the two genes (e.g., genes g_x and g_y) is indeed induced because of a mutual association with the regulating molecules (e.g., miR-5p/-3p pair), the observed degree of coexpression would be expected to be lower after the regulating molecules are removed. In such a scenario, the degree of reduction of gene coexpression may quantify how strong the association is between the regulators and the targets, and this could be utilized as a feature in predicting miR-5p/-3p coregulated target genes.

Pearson product moment or Spearman's rank-correlation coefficient (Spearman's ρ) only computes the expression correlation between g_x and g_y , and confounds our ability to ascertain the effect of common regulating molecule(s) in the observed gene coexpression event (Barzel and Barabasi 2013). However, this limitation can be overcome by the use of partial correlation statistic. By definition, a partial correlation measures gene coexpression with the effect of a set of controlling variables removed (Crawley 2005; Chen and Zheng 2009). In our study, the two mature miRNA arms are the two controlling variables; hence, we used second-order partial correlation (SOPC). We computed the reduction of gene coexpression in terms of "relative difference" between the Spearman's ρ and the SOPC, and then carefully utilized that in predicting miR-5p/-3p coregulated target genes. Next, we computed the gene expression correlation for each pair of interacting proteins in the curated human protein-protein interaction (PPI) network (Cowley et al. 2012) to find a meaningful biological context for miR-5p/-3p coregulated genes.

This study elucidated several novel and important insights. Firstly, our computational approach allowed us to systematically explore the miR-5p/-3p coregulation ability in cancer. Secondly, we applied our approach to lung cancer data and identified a large pool of aberrantly expressed miR-5p/-3p pairs. Thirdly, we derived a new scoring system that prioritized dysregulated protein complexes in lung cancer predicted to be targeted by the aberrantly expressed miR-5p/-3p

pairs. Finally, a network functional analysis revealed a candidate miR-5p/-3p pair that may induce uncontrolled cell proliferation and alter the cell cycle progression process in lung cancer by modulating the expression of genes in the FOXM1 transcription factor network pathway. These results demonstrate that our network-assisted analyses could elucidate potentially critical miR-5p/-3p pairs, as well as new insights into the regulatory mechanisms in lung cancer.

RESULTS

Concordantly dysregulated miR-5p/-3p arms in lung cancer

We started our analyses by using a large pool of 243 lung squamous cell carcinoma (LUSC) patient samples and 36 normal lung tissue samples to pinpoint differentially expressed (DE) miR-5p/-3p pairs in lung cancer. The miRNA expression data were obtained from The Cancer Genome Atlas (TCGA) data portal (see Materials and Methods). Relevant clinical characteristics of the available patient samples are provided in Supplemental Table S1.

It has been observed that subtle changes in miRNA expression; for example, a 1.5-fold change, can have significant impacts on the biology of the cell (Mestdagh et al. 2009; Gorter et al. 2014). By combining a 1.5-fold change with a strictly adjusted P value cut-off of $<10^{-3}$ (adjusted using the Benjamini-Hochberg (BH) multiple testing correction method (Benjamini and Hochberg 1995), we obtained 23 miR-5p/-3p pairs that were DE in LUSC compared with normal lung tissue samples (see Materials and Methods and Table 1). Remarkably, the two DE arms from the same pre-miRNA showed concordant dysregulation pattern; that is, both arms were either significantly up-regulated or down-regulated in LUSC (Fig. 1), and they were highly coexpressed (Spearman's ρ ranged from 0.3 to 0.9, median = 0.78, correlation adjusted $P < 10^{-6}$, adjusted by the BH method) (Supplemental Fig. S2).

Of note, we also analyzed four other cancer types including breast invasive carcinoma, lung adenocarcinoma, kidney renal clear cell carcinoma, and prostate adenocarcinoma using large-scale miRNA-seq expression profiles extracted from TCGA-data portal (sample information is provided in Supplemental Table S2). Intriguingly, we found the same dysregulation pattern as observed in Figure 1 (Supplemental Fig. S3). Existence of such a large pool of concordantly dysregulated miR-5p/-3p pairs, in each cancer, indicates that collectively they may play important role in tumorigenesis.

An integrative framework for miR-5p/-3p coregulatory network construction

Here, we constructed a computational framework to predict miR-5p/-3p coregulated protein interaction networks in lung cancer by integrating concordantly dysregulated miR-5p/-3p

TABLE 1. List of 23 miRNAs whose both arms were differentially expressed in LUSC compared with normal lung tissue

miRNA	Expression pattern	-5p arm		-3p arm	
		Fold change	Adjusted <i>P</i> value ^a	Fold change	Adjusted <i>P</i> value ^a
miR-101	Down	-2.64	9.34×10^{-36}	-3.55	5.44×10^{-58}
miR-126	Down	-3.72	2.23×10^{-73}	-2.89	6.62×10^{-30}
miR-139	Down	-4.31	8.34×10^{-42}	-4.49	5.19×10^{-39}
miR-140	Down	-1.52	1.38×10^{-10}	-2.92	2.19×10^{-41}
miR-144	Down	-8.89	8.01×10^{-63}	-5.63	1.86×10^{-33}
miR-145	Down	-1.97	3.05×10^{-15}	-2.40	4.36×10^{-25}
miR-181a	Down	-2.45	1.26×10^{-28}	-1.94	3.21×10^{-16}
miR-195	Down	-2.91	9.79×10^{-38}	-2.46	1.90×10^{-29}
miR-29c	Down	-1.56	5.00×10^{-8}	-2.26	2.98×10^{-13}
miR-3065	Down	-1.71	1.67×10^{-4}	-3.55	5.67×10^{-24}
miR-30a	Down	-7.86	2.42×10^{-108}	-5.71	5.56×10^{-60}
miR-30c	Down	-1.58	1.51×10^{-8}	-5.08	4.16×10^{-54}
miR-338	Down	-6.61	1.40×10^{-64}	-4.72	1.77×10^{-34}
miR-106b	Up	1.62	1.46×10^{-7}	2.12	2.56×10^{-12}
miR-130b	Up	7.58	8.24×10^{-50}	6.08	2.41×10^{-42}
miR-141	Up	3.24	4.77×10^{-21}	3.47	3.30×10^{-22}
miR-19b	Up	1.65	8.26×10^{-5}	1.63	1.22×10^{-5}
miR-200a	Up	3.90	1.36×10^{-23}	2.46	4.11×10^{-11}
miR-200c	Up	3.04	1.01×10^{-19}	2.51	1.26×10^{-19}
miR-21	Up	1.84	6.48×10^{-11}	1.90	6.26×10^{-10}
miR-324	Up	2.37	3.28×10^{-15}	2.15	3.25×10^{-12}
miR-708	Up	11.39	9.10×10^{-48}	8.32	3.32×10^{-39}
miR-93	Up	2.36	1.42×10^{-14}	3.69	1.14×10^{-31}

^a*P* values were adjusted using the Benjamini and Hochberg (1995) multiple testing correction method.

pairs in LUSC, matched miRNA and mRNA expression profiles in LUSC, and the gene coexpression after mapping it to the human PPI network (see Materials and Methods; Fig. 2A–C). We integrated human PPI data and mRNA-seq expression profiles by computing the expression correlation of the PPI pairs to construct a comprehensive and context-specific Coexpressed PPI (CePPI) network. This resulted in 20,427 CePPI pairs in which the observed gene expression correlations were significant (Spearman's $\rho > 0.3$, adjusted $P < 10^{-6}$, adjusted by the BH method) (see Materials and Methods; Fig. 2B). We applied sequence-based and expression profile-based miRNA-target identification features (Grimson et al. 2007; Bandyopadhyay and Mitra 2009; Liu et al. 2014) to predict miR-5p/-3p coregulated CePPI pairs, termed as CoRegulation (CoReg) motifs (see Materials and Methods; Fig. 2C). To minimize false-positive results, we filtered the CoReg motifs by using the following procedure: For each potential CoReg motif, we computed the second-order partial correlation (SOPC) (Crawley 2005; Chen and Zheng 2009) to infer the gene expression correlation of the CePPI pair after removing the effect of the potential coregulating miR-5p/-3p pair (see Materials and Methods). Subsequently, we computed the “relative difference” between the Spearman's ρ and SOPC to measure changes in gene coexpression. The “relative difference” values were then converted to Z-scores, which we termed as coregulation motif (Z_{CoReg}) scores. The Z_{CoReg} score may explain the strength

of the regulator–target association in a CoReg motif. The obtained Z_{CoReg} scores were subsequently used to generate *P* values by comparing them with the normal distribution (see Materials and Methods). We selected 2356 potential

FIGURE 1. Concordant dysregulation of miR-5p and miR-3p arms in LUSC. miR-5p/-3p pairs are denoted by empty circles if both arms were up- and down-regulated, respectively, in tumor samples compared with normal tissue samples. Solid circles denote the miR-5p/-3p pairs whose one or both arms were not differentially expressed. The x-axis and y-axis represent the expression fold change of the miR-3p and miR-5p arm, respectively.

FIGURE 2. Computational framework for identifying miR-5p/-3p mediated cooperative regulations in lung cancer. The computational framework has the following four steps. (A) Identification of significantly differentially expressed (DE) miR-5p/-3p pairs. (B) Integration of gene coexpression and human protein–protein interaction (PPI) network to obtain significantly coexpressed PPI (CePPI) pairs. (C) Each CePPI pair has two genes g_x and g_y at the transcript level. In the 3' UTR of g_x or g_y , potentially effective miRNA binding sites (6mer/7mer-A1/7mer-m8/8mer) were identified. If miR-5p (or miR-3p) arm potentially targets g_x or g_y , and the miRNA–gene pair was inversely correlated, the pair was selected. Subsequently, miR-5p/-3p-mediated coregulation (Z_{CoReg}) score was derived to select miR-5p/-3p-mediated CoRegulation (CoReg) motifs with a Z_{CoReg} threshold of >1.6 and $P < 0.05$ (see Materials and Methods). Red and green triangles represent up- and down-regulated miRNAs, respectively, in lung cancer. (D) Gene coexpressions become significantly lower ($P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test) with the effect of potentially regulating miR-5p/-3p arms removed. The effect was removed by using the second-order partial correlation statistic. The reduction of gene coexpression is shown in the box plots (inset) along with the median (line within the box), 25th and 75th percentiles, and whiskers with $1.5 \times$ the interquartile range.

CoReg motifs by applying a Z_{CoReg} cut-off value >1.6 and $P < 0.05$.

For the selected CoReg motifs, consecutive measuring of Spearman's ρ (median=0.43) and SOPC (median=0.28) pinpointed significant reduction of expression correlation of the CePPI pairs ($P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test). Of note, we observed that the median SOPC score was lower than the smallest Spearman's ρ score (Fig. 2D). The selected CoReg motifs consisted of 17 miR-5p/-3p pairs and 1618 CePPI pairs. The number of predicted targets for each miR-5p/-3p pair is summarized in Supplemental Figure S4.

Subsequently, we computed the first-order partial correlation (FOPC). An FOPC computes gene expression correlation after removing the effect of a single controlling variable (here, either miR-5p or miR-3p arm). The FOPC measurement is required to examine the effect of an individual miRNA arm

on the expression correlation of the potentially targeted CePPI pairs. We observed that gene coexpressions were reduced significantly with the effect of the individual miRNA arm was removed (Fig. 3). These results may suggest that gene coexpression was strongly and unbiasedly influenced by both miR-5p and miR-3p arms in the selected CoReg motifs.

Prioritizing miR-5p/-3p coregulated protein complexes in lung cancer

Considering that the functional units in a cellular system are often protein complexes rather than individual proteins (Sass et al. 2011), we prioritized miR-5p/-3p coregulated protein complexes to decipher the functional relevance of the two mature miRNA arms that originated from the same

FIGURE 3. Reduction of gene coexpression with the effect of miR-5p or miR-3p arm removed. For each miR-5p/-3p pair, predicted CoReg motifs were selected. The analysis shows comparative results of correlation (measured by Spearman's ρ), first-order partial correlation (FOPC) with the effect of miR-5p arm removed, and FOPC with the effect of miR-3p arm removed. Box plots elucidate the reduction of gene coexpression after removing the effects of only the miR-5p arm or only the miR-3p arm, compared with the correlation value. The inside table shows the statistical significance of the reduction of gene coexpression (Wilcoxon rank-sum test). Results for the miR-181a-5p/-3p pair are not shown because the number of CoReg motifs=2.

precursor product. Among the 20,427 background CePPI pairs identified above, 1618 belonged to the selected CoReg motifs. We first identified hub proteins from the background CePPI networks. A protein was labeled a hub if the protein had at least 10 interacting protein partners (Haynes et al. 2006). A degree cut-off value of 10 corresponded to approximately the top 15% of the node degree distribution. We derived protein complexes by selecting the hub's direct interacting partners in the network, as described by Börnigen et al. (2013) (see Materials and Methods). We obtained 1041 protein complexes, among which 396 were linked with at least one miR-5p/-3p pair.

We introduced a scoring system to prioritize the protein complexes that had biologically meaningful associations with the miR-5p/-3p pairs. The scoring system selected 35 protein complexes that were significantly enriched with the miR-5p/-3p coregulated CePPI pairs (Hypergeometric test, adjusted $P < 10^{-3}$, adjusted by the BH method). Subsequently, we ranked the complexes based on the derived scores, termed as composite score (CS). The CS takes into account the average expression correlation of the complex members, the average Z_{CoReg} score, and the average inverse correlation between the miRNA and potential targets in the complex (see Materials and Methods for details). These three

attributes are biologically meaningful. The first term reflects how strong the association is between the complex members. The second term explains the influence of potentially regulating miR-5p/-3p pairs in the association of the complex members. Finally, the third term reflects the strength of the inverse association between the miRNA and potential target pair. Complex-wise predicted CoReg motifs are listed in Supplemental Table S3.

Among the 35 complexes, 21 were potential targets of the eight up-regulated miR-5p/-3p pairs in LUSC. We categorized these complexes as Group 1. The remaining 14 complexes were categorized as Group 2; they were potential targets of nine down-regulated miR-5p/-3p pairs (Fig. 4). For the Group 1 complexes, predicted targets were down-regulated in LUSC compared with normal lung tissue samples. Intriguingly, among the 21 complexes, the majority (18 or 85.71%) had a median down-regulation of more than twofold. For the Group 2 complexes, the predicted targets were up-regulated. Approximately half of these complexes (8 out of 14 or 57.14%) had a median up-regulation of more than twofold. The up- or down-regulation of protein complexes was synchronized with the down- or up-regulation of miR-5p/-3p pairs, respectively, which is biologically meaningful. These results support the general understanding

of miRNA-mediated target repression. Of note, we did not consider differential gene expression values to derive the CS; hence, the observed associations were not due to the underlying scoring strategy.

For the top 10 complexes, we conducted a functional analysis using the biological process terms of Gene Ontology (GO) (Ashburner et al. 2000) by using the software WebGestalt (Zhang et al. 2005). For each complex, we selected only the top five significantly over-represented biological processes (hypergeometric test, adjusted $P < 0.05$, adjusted by the BH method) using the tool REVIGO, which summarizes the list of nonredundant GO terms (Supek et al. 2011). We identified all the complexes associated with at least one GO term that is closely related to the cancer-associated biological process, such as “regulation of apoptotic process,” “regulation of cell differentiation,” and “cell cycle checkpoint” (Supplemental Fig. S5). The results suggest that the aberrant expression of these complexes may interrupt the normal biological functions and contribute tumorigenesis.

Next, we selected one candidate miR-5p/-3p pair and performed an in-depth analysis using a systems biology approach to determine its pathogenic potential in lung cancer. Among the 17 miR-5p/-3p pairs, miR-145-5p and miR-145-3p were found to be not only down-regulated in LUSC (Table 1), but

were also down-regulated in highly aggressive osteosarcoma cell lines compared with nonaggressive cell lines (5.1- and 4.2-fold down-regulation, q -value = 0.014 and 0.016, respectively) (Lauvrak et al. 2013), as well as in intracranial aneurysm patients compared with controls (more than twofold down-regulation, $P < 0.05$; down-regulation was further confirmed by qRT-PCR) (Jiang et al. 2013). While multiple studies (Jiang et al. 2013; Lauvrak et al. 2013), including ours, reported the down-regulation of this miRNA pair in different diseased conditions, little is known about its coregulatory mechanism in lung cancer or other diseases.

miR-145-5p/-3p pair potentially alters ‘cell cycle progression,’ leading to uncontrolled cell proliferation

From the selected 2356 CoReg motifs, we obtained 325 protein-coding genes that were predicted to be coregulated by the miR-145-5p/-3p pair. For each gene, we computed the average Spearman’s ρ from miR-145-5p-gene pairs and miR-145-3p-gene pairs. The top 20% (65 genes) highly anti-correlated genes (ranging from -0.48 to -0.59) were selected

FIGURE 4. Protein complexes predicted to be coregulated by miR-5p/-3p pairs. The x-axis represents 35 protein complexes that were potentially coregulated by miR-5p/-3p pairs. The complex name is represented by the hub protein followed by the rank (example: TGFBR2_1 represents a protein complex in which TGFBR2 is the hub protein, and the rank of this complex is 1). Two distinct groups of complexes were observed. Members of 21 complexes showed down-regulation, while the remaining 14 complexes showed up-regulation. Down- or up-regulated complexes were predicted to be targeted by one or multiple up- (heat map; values denote the fold change) or down-regulated (heat map; values denote the fold change) miRNAs. The y-axis represents a log2 fold change of complex members that were predicted to be targeted by the miR-5p/-3p pair. The median fold change is shown by the line within the box.

for pathway enrichment analysis. We selected the top 10 significantly enriched pathways (BH adjusted $P < 10^{-9}$; Supplemental Table S4) from the Pathway Common database (Cerami et al. 2011), which is embedded into the software WebGestalt (Zhang et al. 2005). These pathways are associated with functions related to “cell cycle checkpoints,” “cell cycle stages,” and “cell cycle progression.” One of the top 10 pathways was the “FOXM1 transcription factor network,” which consisted of a transcription factor (TF) FOXM1 and five genes: *BIRC5*, *CCNA2*, *CCNB1*, *CKS1B*, and *NEK2*. Literature-based evidence suggests that FOXM1 directly regulates these genes (Wonsey and Follettie 2005; Zhan et al. 2007; Petrovic et al. 2008; Li et al. 2012). Importantly, both FOXM1 and these genes were up-regulated (ranging from 4.96- to 24.93-fold) in LUSC (Supplemental Table S5) and their elevated expression alters cell cycle progression, leading to uncontrolled cell proliferation, which is the hallmark of cancer. Another proliferation-related gene *CCNE2* (6.32-fold up-regulated in LUSC; belongs to the top 20% anti-correlated genes, as mentioned above), is also involved in cell cycle progression (Gudas et al. 1999). Hence, we included the TF FOXM1 and the six genes, including *CCNE2*, to construct a miR-145-5p/-3p coregulatory sub-network specific to the process of “cell cycle progression” (Fig. 5A). The results suggest that the elevated expression of these genes and the TF FOXM1 could be the consequence of significant down-regulation of miR-145-5p/-3p pair in LUSC. Importantly, we reexamined miR-145-5p/-3p-FOXM1, miR-145-5p/-3p-CCNB1, and miR-145-5p/-3p-CCNE2 associations using a cohort of 104 matched miRNA and protein expression profiles in LUSC to determine whether the regulatory relationships were consistent at both the transcriptional and protein levels.

miR-145-5p/-3p pair potentially modulates FOXM1, CCNB1, and CCNE2 protein expression in a synergistic manner

Among the 243 LUSC patient samples that were used in this study, 104 samples also had reverse-phase protein array (RPPA) data for FOXM1, CCNB1, and CCNE2 proteins (a subset of the predicted targets mentioned in Fig. 5A) in the Cancer Proteome Atlas (TCPA) database (Li et al. 2013), which is part of the TCGA project. We sorted the 104 patient samples in ascending order by miR-145-5p expression values and divided them into four quartile groups. Separately, we performed the same task for miR-145-3p. If the samples belonged to the first 25th percentile in both miR-145-5p and miR-145-3p expression profiles, we denoted these samples as “both low.” If the samples belonged to the first 25th percentile in one expression profile, we denoted these samples as “one low.” Similarly, we denoted the samples as “both high” and “one high” if they were in the 75th–100th percentile range in both expression profiles and one expression profile, respectively. The number of samples and the patient IDs in each category are provided in Supplemental Table S6.

We observed that FOXM1, CCNB1, and CCNE2 protein expressions were significantly elevated in the samples categorized as “both low” ($P = 2.52 \times 10^{-4}$, 3.29×10^{-4} , and 0.01 for FOXM1, CCNB1, and CCNE2, respectively), compared with the samples designated as “both high,” or “one low” ($P = 0.02$, 0.02, and 0.09, respectively; the last P -value is marginally significant, but showed the same trend). Similarly, elevated expressions were also observed when we compared “both low” with “one high” group of samples (Fig. 5B). These results suggest that simultaneous down-regulation of miR-145-5p/-3p

FIGURE 5. miR-145-5p/-3p pair potentially alters the “cell cycle progression” process in lung cancer. (A) “FOXMI transcription factor network” pathway members were predicted to be coregulated by the miR-145-5p/-3p pair. Along with the gene *CCNE2*, these pathway molecules were significantly up-regulated in LUSC, leading to the altered “cell cycle progression” process. Triangle nodes: down-regulated miRNA; square node: up-regulated TF; circle nodes: up-regulated genes. Up- or down-regulation was measured in LUSC compared with normal lung tissue samples. (B) Protein expression of the predicted targets was modulated by coregulating miR-5p/-3p expression. The samples have matched miRNA and protein expression profiles. Sample grouping was performed based on the expression levels of miR-145-5p and miR-145-3p. Both low: miR-145-5p and miR-145-3p expressions were low (belonging to the first 25th percentile after sorting the samples by expression level). One low: either miR-145-5p or miR-145-3p expression belonged to the first 25th percentile. Both high: miR-145-5p and miR-145-3p expressions were high (belonging to the 75th–100th percentile). One high: either miR-145-5p or miR-145-3p expression belonged to the 75th–100th percentile. Statistical significance in expression difference was marked as (***) $P < 0.001$, (**) $P < 0.05$, and (*) $P < 0.1$ (Wilcoxon rank-sum test).

pair might be the reason for the elevated protein expression of FOXM1, CCNB1, and CCNE2. Using this data set, we further found a strong inverse association between miR-145-5p-FOXM1 (Spearman $\rho = -0.27$, $P = 2.56 \times 10^{-3}$), miR-145-3p-FOXM1 (Spearman $\rho = -0.36$, $P = 7.66 \times 10^{-5}$), miR-145-5p-CCNB1 (Spearman $\rho = -0.32$, $P = 4.25 \times 10^{-4}$), miR-145-3p-CCNB1 (Spearman $\rho = -0.34$, $P = 2.01 \times 10^{-4}$), miR-145-5p-CCNE2 (Spearman $\rho = -0.32$, $P = 4.22 \times 10^{-4}$), and miR-145-3p-CCNE2 (Spearman $\rho = -0.22$, $P = 1.10 \times 10^{-2}$). In summary, these results are reproducible in large-scale transcriptomic and proteomic data. The reproducible regulatory relationships likely reflect the true biological regulations in a cellular system, though future studies of their functions are warranted.

DISCUSSION

In this study, for the first time, we constructed miR-5p/-3p coregulatory protein interaction networks in lung cancer (LUSC). Our framework started with the identification of concordantly dysregulated miR-5p/-3p pairs in LUSC. Next, we integrated gene expression correlation onto the human PPI network to generate a meaningful biological context in terms of functional association for miR-5p/-3p coregulated genes. One important outcome of this comprehensive study was the identification of 17 concordantly dysregulated miR-5p/-3p pairs in LUSC that potentially modulate the expression of cancer-associated protein complexes. This suggests that the reported miR-5p/-3p pairs may have critical roles in tumor development and/or progression.

Furthermore, we found notably strong inverse associations between miR-145-5p/-3p-FOXM1, miR-145-5p/-3p-CCNB1, and miR-145-5p/-3p-CCNE2 pairs in both the large-scale transcriptional and protein expression profiling data that have matched miRNA expression profiles. These reproducible molecular associations within the miR-145-5p/-3p coregulatory sub-network increase the confidence in the miR-145-5p/-3p-mediated alteration of the “cell cycle progression” process, which may ultimately result in uncontrolled cell proliferation.

One main issue in this large-scale computational data analysis is to control the false positives during the construction of regulatory relationships. To minimize the effect of false positives, an additional level of filtering was introduced by using a novel miRNA-target identification feature called “relative difference,” which was designed based on the premise that coregulating molecule(s) may induce the coexpression of their common target genes (Barzel and Barabasi 2013). It is imperative to mention that we evaluated the predictive power of “relative difference” using the experimentally validated miRNA-target-gene pairs extracted from the TarBase 6.0 database (Vergoulis et al. 2012). We identified 257 highly confident CoReg motifs (Supplemental Table S7), of which 254 (98.83%) showed a prominent reduction in expression correlation ($P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test) of the ex-

perimentally validated target-gene pairs after removing the effect of coregulating miRNA pairs (Supplemental Fig. S6). Furthermore, we compiled 27 non-CoReg motifs, each of which contains at least one miRNA-nontarget-gene pair from experimental data (reporter gene assay). These miRNA-nontarget-gene pairs were detected as false-positives (Supplemental Table S8) after using sequence-based and expression profiling-based (anti-correlation) filtering steps mentioned in Figure 2C. The false-positive target genes were then used to extract coexpressed PPI (CePPI) pairs. We then constructed 27 non-CoReg motifs (listed in Supplemental Table S9) by linking the CePPI pairs with the corresponding miR-5p/-3p arms, and in this case one interaction in a motif is a known false positive. After removing the effect of the miR-5p/-3p arms using the second-order partial correlation, the result showed either no reduction or very weak reduction of gene coexpression (Supplemental Fig. S7). This evaluation result suggests that the feature “relative difference” provides an additional level of filtering to remove miR-5p/-3p coregulated false-positive motifs.

There are some limitations in this study. First, due to the unavailability of independent miR-5p and miR-3p expression profiling data in LUSC, we were unable to examine whether the concordant dysregulation patterns of miR-5p/-3p pairs are reproducible. However, we observed that the miR-145-5p/-3p pair is down-regulated in multiple diseased conditions (Jiang et al. 2013; Lauvrak et al. 2013), which supports our results. Second, the expression profiles (mRNA, miRNA, and protein) we used in this study may not truly reflect the dynamic regulation in a cellular system, especially considering that cancer is highly heterogeneous. However, our integrative approach likely allows us to overcome such a limitation and identify the major regulation patterns.

In summary, we demonstrated that miR-5p/-3p coregulation is an important regulatory mechanism in developing cancer. The results provide important candidates for further experimental validation and, if validated, may help us to have a deeper understanding of the regulatory mechanisms of lung cancer.

MATERIALS AND METHODS

Analysis of mRNA and miRNA expression profiles

We obtained RNA-seqV2 data for mRNA expression and miRNA-seq data for miRNA expression from 243 tumor and 36 normal lung tissue samples, available in TCGA-data portal (<https://tcga-data.nci.nih.gov/tcga/>). The data sets consist of both read counts and normalized expression values for mRNAs and mature miRNAs.

We used the read counts to identify differentially expressed genes and miRNAs. The read counts were imported into the R/Bioconductor package edgeR (Robinson et al. 2010). The differential expression of genes/miRNAs between tumor and normal samples was assessed by estimating an exact test P value, similar to Fisher's exact test. The results were further adjusted using the

Benjamini–Hochberg (BH) multiple testing correction method (Benjamini and Hochberg 1995).

To compute the Spearman rank-correlation (Spearman's ρ) and partial correlation coefficient, we used normalized mRNA and miRNA expression profiles. The R package `cor.test` and `ppcor` (<http://CRAN.R-project.org/package=ppcor>) were used to compute the correlation and partial correlation, respectively.

Coexpressed protein–protein interaction networks

PPIs were downloaded from the Protein Interaction Network Analysis version 2 (PINA v2.0) database, which collected and curated 108,477 PPI pairs from six manually curated public repositories (Cowley et al. 2012). Only the experimentally validated PPIs were used. For each PPI pair, we computed Spearman's ρ using the mRNA expression profiles of 243 tumor and 36 normal lung tissue samples (Fig. 2B). Finally, we obtained 20,427 significantly coexpressed PPI (CePPI) pairs (Spearman's $\rho > 0.3$, adjusted $P < 10^{-6}$, adjusted by the BH method).

Prediction of miRNA-mediated gene repression

We obtained 4138 protein-coding genes from 20,427 CePPI pairs. For these genes, we downloaded their 3' untranslated region (UTR) annotations from the University of California, Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/>). For human mature miRNA sequence, we downloaded the data from the miRBase database (Release 20) (Kozomara and Griffiths-Jones 2014). miRNA-target relationships were predicted by using miRNA-target identification features, as explained in Bandyopadhyay and Mitra (2009). We first identified miRNA binding sites in the 3' UTR of the target transcript. The binding sites should be of one of the 6mer/7mer-A1/7mer-m8/8mer (Bandyopadhyay and Mitra 2009). A predicted miRNA-target-gene pair must have at least one potentially effective or functional miRNA binding site. Binding sites were defined as potentially effective if the sites do not reside too close (≤ 15 nt) to the stop codon nor near the middle of the 3' UTR of the target transcript (Grimson et al. 2007; Bandyopadhyay and Mitra 2009). We divided the entire 3' UTR sequence of a transcript into ten bins, and a site was regarded as effective if it resided in one of the four terminal bins from both ends (Fig. 2C; Bandyopadhyay and Mitra 2009). Thereafter, for each miRNA-gene pair, we computed Spearman's ρ using the matched miRNA and mRNA expression profiles. Significantly anti-correlated (BH adjusted $P < 0.05$) miRNA-gene pairs were selected for follow-up analysis.

Prediction of CoReg motifs using second-order partial correlation

We applied partial correlation to remove the effect of miR-5p and miR-3p arms in the gene expression correlation of CePPI pairs. A partial correlation measures the degree of association between two random variables with the effect of controlling random variable(s) removed (Crawley 2005; Chen and Zheng 2009). The order of partial correlation is determined by the number of controlling random variables. If the controlling variable is one, we compute the first-order partial correlation (FOPC). Since miR-5p and miR-3p are the

two controlling variables, we computed the second-order partial correlation coefficient (SOPC).

Let us assume that, at the transcript level, a CePPI pair has two genes, g_x and g_y . Let us also assume that 5p and 3p represent the miR-5p arm and miR-3p arm, respectively. The Spearman's ρ is denoted by $r_{g_x g_y}$. The FOPC between g_x and g_y conditioning on 5p is

$$r_{g_x g_y, 5p} = \frac{r_{g_x g_y} - r_{g_x 5p} r_{g_y 5p}}{\sqrt{(1 - r_{g_x 5p}^2)(1 - r_{g_y 5p}^2)}}.$$

The SOPC between g_x and g_y conditioning on 5p and 3p is

$$r_{g_x g_y, 5p 3p} = \frac{r_{g_x g_y, 5p} - r_{g_x 3p, 5p} r_{g_y 3p, 5p}}{\sqrt{(1 - r_{g_x 3p, 5p}^2)(1 - r_{g_y 3p, 5p}^2)}}.$$

Therefore, for each CePPI pair, we obtained correlation ($r_{g_x g_y}$) and SOPC ($r_{g_x g_y, 5p 3p}$). If $r_{g_x g_y, 5p 3p} < r_{g_x g_y}$, we computed the “relative difference” (d_r) between the correlation and the partial correlation coefficient by

$$d_r = \frac{r_{g_x g_y} - r_{g_x g_y, 5p 3p}}{((r_{g_x g_y} + r_{g_x g_y, 5p 3p})/2)}.$$

From the list of d_r values, we computed the Z-scores (termed as Z_{CoReg}) by

$$Z_{\text{CoReg}} = \frac{d_r - \mu}{\sigma}.$$

We then transformed the computed Z_{CoReg} scores into P values by comparing with normal distribution. The motifs having a cut-off value > 1.6 and $P < 0.05$ were selected as potential CoReg motifs.

Coordinated expression of protein complexes

Let a protein complex C consists of a hub protein H and its interacting partner proteins $I_1, I_2, \dots, I_n$, $n > 10$. We computed the gene expression correlation (Spearman's ρ) between H and I_i (denoted by $r_{I_i, H}$) across 279 samples (243 LUSC and 36 normal lung tissue samples). Average of these single coexpression values (AC_C) represent the “coordinated expression” of complex C (Börnigen et al. 2013):

$$AC_C = \sum_{i=1}^n \frac{r_{I_i, H}}{n}.$$

The “coordinated expression” explains how strongly the member proteins in a complex are connected to the hub protein (Börnigen et al. 2013).

Prioritizing miR-5p/-3p coregulated protein complexes

To assess if a miR-5p/-3p pair cooperatively regulates a significant portion of the members in a protein complex C , we performed an enrichment analysis based on the hypergeometric test and computed the P value (P_C) as below:

$$P_C = \sum_{i=k}^n \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}},$$

where N is the total number of CePPI pairs; m is the number of CePPI pairs predicted to be coregulated by a miR-5p/-3p pair

with $Z_{\text{CoReg}} > 1.6$ and $P < 0.05$, n is the number of $r_{i,H}$ pairs, $i = 1, 2, \dots, n$, in the protein complex C , and k is the number of $r_{i,H}$ pairs in the protein complex C with $Z_{\text{CoReg}} > 1.6$ and $P < 0.05$. The obtained P values were further corrected for multiple testing using the BH method. All protein complexes with an adjusted $P < 10^{-3}$ were selected. We ranked the selected complexes by deriving the composite scores (CS).

A composite score for protein complex C (CS_C) was computed by integrating three values: AC_C , average $Z_{\text{CoReg}}(AZ_{\text{CoReg}C})$, and the average inverse correlation (AIC_C) between the miRNA and potential target genes. $\text{Adj.}P_C$ returns a binary score of 1 or 0, depending on whether or not the enriched portion of the complex members was predicted to be coregulated by miR-5p/-3p pairs:

$$CS_C = \text{Adj.}P_C^* \prod (AC_C, AZ_{\text{CoReg}C}, AIC_C),$$

where

$$\text{Adj.}P_C = \begin{cases} 1 & \text{if } \text{Adj.}P_C < 10^{-3}, \\ 0 & \text{otherwise.} \end{cases}$$

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was partially supported by National Institutes of Health (NIH) grants (R01LM011177, R03CA167695, P30CA68485, P50CA095103, and P50CA090949) and Ingram Professorship Funds (to Z.Z.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Ms. Christen Parzych for critically reading and improving an earlier draft of the manuscript.

Received September 16, 2014; accepted February 11, 2015.

REFERENCES

- Amar D, Safer H, Shamir R. 2013. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol* **9**: e1002955.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* **25**: 25–29.
- Bandyopadhyay S, Mitra R. 2019. TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics* **25**: 2625–2631.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Barzel B, Barabasi AL. 2013. Network link prediction by global silencing of indirect correlations. *Nat Biotechnol* **31**: 720–725.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B Met* **57**: 289–300.
- Börnigen D, Pers TH, Thorrez L, Huttenhower C, Moreau Y, Brunak S. 2013. Concordance of gene expression in human protein complexes reveals tissue specificity and pathology. *Nucleic Acids Res* **41**: e171.
- Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. 2011. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* **39**: D685–D690.
- Chang KW, Kao SY, Wu YH, Tsai MM, Tu HF, Liu CJ, Lui MT, Lin SC. 2013. Passenger strand miRNA *miR-31** regulates the phenotypes of oral cancer cells by targeting RhoA. *Oral Oncol* **49**: 27–33.
- Chen L, Zheng S. 2009. Studying alternative splicing regulatory networks through partial correlation analysis. *Genome Biol* **10**: R3.
- Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, Biankin AV, Hautaniemi S, Wu J. 2012. PINA v2.0: mining interactome modules. *Nucleic Acids Res* **40**: D862–D865.
- Crawley MJ. 2005. *Statistics: an introduction using R*. John Wiley and Sons, Chichester, England.
- Gorter JA, Iyer A, White I, Colzi A, van Vliet EA, Sisodiya S, Aronica E. 2014. Hippocampal subregion-specific microRNA expression during epileptogenesis in experimental temporal lobe epilepsy. *Neurobiol Dis* **62**: 508–520.
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91–105.
- Gudas JM, Payton M, Thukral S, Chen E, Bass M, Robinson MO, Coats S. 1999. Cyclin E2, a novel G₁ cyclin that binds Cdk2 and is aberrantly expressed in human cancers. *Mol Cell Biol* **19**: 612–622.
- Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. 2006. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* **2**: e100.
- Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929–934.
- Jiang Y, Zhang M, He H, Chen J, Zeng H, Li J, Duan R. 2013. MicroRNA/mRNA profiling and regulatory network of intracranial aneurysm. *BMC Med Genomics* **6**: 36.
- Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68–D73.
- Lauvrak SU, Munthe E, Kresse SH, Stratford EW, Namlos HM, Meza-Zepeda LA, Myklebost O. 2013. Functional characterisation of osteosarcoma cell lines and identification of mRNAs and miRNAs associated with aggressive cancer phenotypes. *Br J Cancer* **109**: 2228–2236.
- Li Y, Zhang S, Huang S. 2012. FoxM1: a potential drug target for glioma. *Future Oncol* **8**: 223–226.
- Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang JY, Broom BM, Verhaak RG, Kane DW, et al. 2013. TCGA: a resource for cancer functional proteomics data. *Nat Methods* **10**: 1046–1047.
- Liu B, Li J, Cairns MJ. 2014. Identifying miRNAs, targets and functions. *Brief Bioinform* **15**: 1–19.
- Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. 2009. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* **10**: R64.
- Petrovic V, Costa RH, Lau LF, Raychaudhuri P, Tyner AL. 2008. FoxM1 regulates growth factor-induced expression of kinase-interacting stathmin (KIS) to promote cell cycle progression. *J Biol Chem* **283**: 453–460.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Sass S, Dietmann S, Burk UC, Brabletz S, Lutter D, Kowarsch A, Mayer KF, Brabletz T, Ruepp A, Theis FJ, et al. 2011. MicroRNAs coordinately regulate protein complexes. *BMC Syst Biol* **5**: 136.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800.
- Uchino K, Takeshita F, Takahashi RU, Kosaka N, Fujiwara K, Naruoka H, Sonoke S, Yano J, Sasaki H, Nozawa S, et al. 2013. Therapeutic effects of microRNA-582-5p and -3p on the inhibition of bladder cancer progression. *Mol Ther* **21**: 610–619.

- Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T, Hatzigeorgiou AG. 2012. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* **40**: D222–D229.
- Wonsey DR, Follettie MT. 2005. Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe. *Cancer Res* **65**: 5181–5189.
- Yang X, Du WW, Li H, Liu F, Khorshidi A, Rutnam ZJ, Yang BB. 2013. Both mature miR-17-5p and passenger strand miR-17-3p target TIMP3 and induce prostate tumor growth and invasion. *Nucleic Acids Res* **41**: 9688–9704.
- Zhan F, Colla S, Wu X, Chen B, Stewart JP, Kuehl WM, Barlogie B, Shaughnessy JD Jr. 2007. *CKS1B*, overexpressed in aggressive disease, regulates multiple myeloma growth and survival through SKP2- and p27^{Kip1}-dependent and -independent mechanisms. *Blood* **109**: 4995–5001.
- Zhang B, Kirov S, Snoddy J. 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* **33**: W741–W748.

RNA

A PUBLICATION OF THE RNA SOCIETY

Concordant dysregulation of miR-5p and miR-3p arms of the same precursor microRNA may be a mechanism in inducing cell proliferation and tumorigenesis: a lung cancer study

Ramkrishna Mitra, Chen-Ching Lin, Christine M. Eischen, et al.

RNA 2015 21: 1055-1065 originally published online April 7, 2015

Access the most recent version at doi:[10.1261/rna.048132.114](https://doi.org/10.1261/rna.048132.114)

Supplemental Material <http://rnajournal.cshlp.org/content/suppl/2015/04/02/rna.048132.114.DC1>

References This article cites 33 articles, 5 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/21/6/1055.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>