Reproducible combinatorial regulatory networks elucidate novel oncogenic microRNAs in non-small cell lung cancer

RAMKRISHNA MITRA,[1] MICK D. EDMONDS,[2] JINGCHUN SUN,[1] MIN ZHAO,[1] HUI YU,[1]
CHRISTINE M. EISCHEN,[2] and ZHONGMING ZHAO[1,3,4,5]

[1]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA
[2]Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA
[3]Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA
[4]Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee 37212, USA
[5]Center for Quantitative Sciences, Vanderbilt University, Nashville, Tennessee 37232, USA

ABSTRACT

While previous studies reported aberrant expression of microRNAs (miRNAs) in non-small cell lung cancer (NSCLC), little is known about which miRNAs play central roles in NSCLC's pathogenesis and its regulatory mechanisms. To address this issue, we presented a robust computational framework that integrated matched miRNA and mRNA expression profiles in NSCLC using feed-forward loops. The network consists of miRNAs, transcription factors (TFs), and their common predicted target genes. To discern the biological meaning of their associations, we introduced the direction of regulation. A network edge validation strategy using three independent NSCLC expression profiling data sets pinpointed reproducible biological regulations. Reproducible regulation, which may reflect the true molecular interaction, has not been applied to miRNA–TF co-regulatory network analyses in cancer or other diseases yet. We revealed eight hub miRNAs that connected to a higher proportion of targets validated by independent data sets. Network analyses showed that these miRNAs might have strong oncogenic characteristics. Furthermore, we identified a novel miRNA–TF co-regulatory module that potentially suppresses the tumor suppressor activity of the TGF-β pathway by targeting a core pathway molecule (TGFBR2). Follow-up experiments showed two miRNAs (miR-9-5p and miR-130b-3p) in this module had increased expression while their target gene *TGFBR2* had decreased expression in a cohort of human NSCLC. Moreover, we demonstrated these two miRNAs directly bind to the 3′ untranslated region of *TGFBR2*. This study enhanced our understanding of miRNA–TF co-regulatory mechanisms in NSCLC. The combined bioinformatics and validation approach we described can be applied to study other types of diseases.

Keywords: feed-forward loop; lung cancer; microRNA; regulatory network; transcription factor

INTRODUCTION

Lung cancer is the second most common cancer among both men and women in the United States and led to an estimated 160,000 deaths in 2013 (Siegel et al. 2013; Henley et al. 2014). Non-small cell lung cancer (NSCLC) accounts for ∼80%–85% of all cases of lung cancer (Dempke et al. 2010). Despite the advent of new drugs and therapeutic regimens, lung cancer prognosis has not changed significantly in the last two decades (Fabbri et al. 2007). Hence, the development of effective treatments for this disease requires rapid innovations of therapeutic strategies. Recently, several studies have demonstrated that therapeutic strategies based on the modulation of microRNA (miRNA) activities are promising, as these small RNAs have the ability to influence cellular behavior (Kota et al. 2009; Jackson and Linsley 2010; Trang et al. 2010).

MicroRNAs (miRNAs) are small endogenous non-coding RNA molecules that are ∼21–23 nucleotides (nt) in length. In animals, a single-stranded miRNA typically binds to the 3′ untranslated regions (UTRs) of the target mRNA, and this results in mRNA degradation or translational repression of protein production. This process is commonly called post-transcriptional regulation (Bartel 2004). Recent studies indicate aberrant expression of miRNAs can contribute to cancer development by promoting the expression of proto-oncogenes or inhibiting the expression of tumor suppressor genes (Bandyopadhyay et al. 2010). Several miRNAs are reported to have aberrant expression in NSCLC (Guan et al. 2012; Vosa et al. 2013); however, researchers have yet to pinpoint which

miRNAs play critical roles in the pathogenesis of NSCLC or the relevant targets of these miRNAs.

The other type of important regulators during cancer progression is transcription factors (TFs), which bind to genes' promoter regions and either induce or repress gene transcription. A major portion of oncogenes and tumor suppressor genes encode TFs (Libermann and Zerbini 2006) that control gene and miRNA expression and signaling pathways in cancer. The expression of these TFs, on the other hand, can be regulated by miRNAs at the post-transcriptional stage. Therefore, miRNA(s) and TF(s) may regulate each other reciprocally and form a feed-back loop (FBL), or both can jointly regulate the expression of target genes and form a feed-forward loop (FFL) (Guo et al. 2010). FFL-based combinatorial regulatory network approaches recently emerged as promising tools to elucidate complex diseases such as schizophrenia (Guo et al. 2010), glioblastoma multiforme (Setty et al. 2012; Sun et al. 2012), ovarian cancer (Zhao et al. 2013), and osteosarcoma (Poos et al. 2013). However, current FFL studies mostly rely on the predicted regulation information, leading to high false positive rates.

To increase the confidence in true and biologically relevant regulations, one may identify regulatory relationships that are consistent or reproducible in multiple independent studies (Langfelder et al. 2011; Dutta et al. 2012). So far, such a strategy has not been applied to miRNA–TF co-regulatory network analyses in cancer or other diseases. However, with the rapid growth in high-throughput expression profiling studies, this strategy has become not only feasible, but also necessary to identify complex gene regulation in cellular systems.

In this study, we aimed to decipher miRNA- and TF-mediated reproducible regulations in NSCLC by utilizing four matched miRNA and gene expression profiling data sets. We constructed the network by integrating in silico target prediction results and one matched miRNA and gene expression profiling data set. We then applied a network edge validation strategy using three other independent data sets, aiming to pinpoint reproducible biological regulations. A sub-network from a well-known pathway in NSCLC, the transforming growth factor-β (TGF-β) signaling pathway, identified two promising oncogenic miRNAs that cooperatively regulate the core pathway molecule TGFBR2. This regulation was supported by our qRT-PCR and luciferase reporter assays.

RESULTS

An integrative framework for reproducible regulatory network construction

Figure 1 illustrates our integrative framework to construct a comprehensive miRNA–TF co-regulatory network in NSCLC. This network consisted of feed-forward regulations among three major components: miRNAs, genes, and TFs. We restricted our analysis of miRNAs and genes that were differentially expressed (DE) in NSCLC compared with normal lung tissue samples (see Supplemental Fig. S1; Fig. 1A). Specifically, to identify the DE molecules, we extracted expression data from a cohort study of 16 NSCLC patient samples and 16 matched normal lung tissue samples from TCGA (Hammerman et al. 2012). The RNA-Seq data for mRNA expression and matched miRNA-Seq data for miRNA expression were available from the same samples, making our analyses more reliable. Hereafter, we refer to this data set as *Discovery_data* (see Materials and Methods). Details of relevant patient characteristics, study references, and profiling platforms for this data set are provided in Supplemental Tables S1, S2. Using the *Discovery_data*, our analyses yielded 30 up- and 27 down-regulated miRNAs, as well as 680 up- and 823 down-regulated genes in NSCLC compared with normal lung tissue samples that passed the strict DE criteria (twofold change and adjusted P-value $< 10^{-4}$; see Materials and Methods). Known human TFs were collected from the TRANSFAC professional database (release 2011.4) (Matys et al. 2006). We identified 49 TFs (hereafter, referred to as active TFs) that either were DE in NSCLC or had enriched numbers of their predicted targets identified as DE (Materials and Methods) (Fig. 1B,C). In this network, there are four types of regulatory relationships: miRNA regulation of gene expression (miRNA to gene), miRNA regulation of TF expression (miRNA to TF), TF regulation of gene expression (TF to gene), and TF regulation of miRNA expression (TF to miRNA). We nominated regulator-mediated target activation or repression event based on the observed positive or negative expression correlation computed by Spearman rank correlation statistic (Fig. 1D). To integrate these regulations into a miRNA–TF co-regulatory network, we only included miRNA- and TF-mediated FFLs (Fig. 1E).

To assess if the observed expression correlations between regulator-target pairs were reproducible, we introduced a network edge validation strategy using three independent NSCLC expression profiling data sets, *Validation_data*1, *Validation_data*2, and *Validation_data*3, among which the matched miRNA and gene expression profiles are available (see Materials and Methods). Relevant patient characteristics, study references, and profiling platforms for these data sets are described in Supplemental Tables S1 and S3–S5. Of note, *Discovery_data* was used to construct the miRNA–TF co-regulatory network and introduce the direction of the biological regulations by measuring the expression correlation, while the three validation data sets were used to assess whether the inferred expression correlation between a regulator and a target could be consistently observed in independent studies. We performed functional enrichment and pathway analyses to pinpoint important regulatory modules, their component miRNAs, and their targets that have pathogenic potential in the NSCLC progression. Follow-up experimental results confirmed the reliability of a regulatory module associated with the important TGF-β pathway (Fig. 1E).
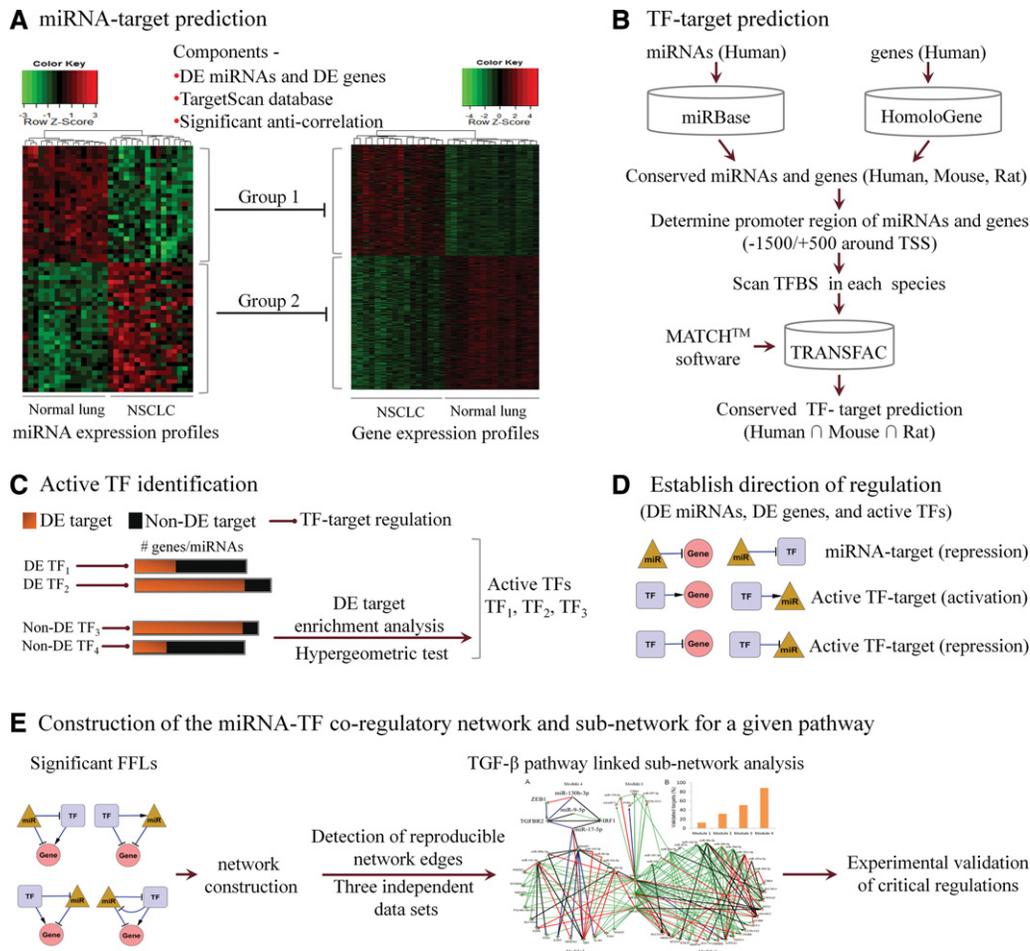
FIGURE 1. Schematic flowchart for constructing and analyzing the comprehensive miRNA–TF co-regulatory networks in NSCLC. miR represents microRNA (miRNA). The computational framework has the following major steps. (*A*) Prediction of miRNA-mediated gene repression based on differentially expressed (DE) miRNAs and genes, target prediction results from TargetScan database, and significant anti-correlation measured by Spearman rank correlation statistic. Group 1 consisted of down-regulated miRNAs and up-regulated genes and Group 2 consisted of up-regulated miRNAs and down-regulated genes. Up- or down-regulation was measured in NSCLC compared with the normal lung tissue. (*B*) A pipeline to identify conserved TF and target gene/miRNA relationships. (*C*) Identification of active TFs. A TF is regarded as active if its encoding gene is DE or its predicted targets are statistically enriched with DE genes/miRNAs in NSCLC compared with the normal lung tissue (hypergeometric test, *P*-value < 0.05; see Materials and Methods). (*D*) Establishment of the direction of regulation by measuring expression correlation between a regulator and a target. (*E*) Construction of the miRNA–TF co-regulatory network and sub-network from a given pathway. The network consisted of 32 DE miRNAs, 31 active TFs, 147 DE genes, and 688 edges. Edges with different colors represent their validation status as found using three independent data sets (details are shown in Fig. 3). To confirm the regulatory relationships in one pathway module, follow-up experiments were conducted using quantitative RT-PCR and luciferase reporter assays.

Highly confident regulatory relationships among miRNAs, TFs, and genes

Table 1 summarizes four types of potential regulatory relationships explained above and related methods. We provided more details below.

For a miRNA, we predicted conserved miRNA to gene relationships by parsing the TargetScan prediction results (Friedman et al. 2009). To assign the direction of the regulation, we combined the prediction results and *Discovery_data*. We predicted miRNA-mediated repression of genes if (1) miRNAs and their predicted targets were discordantly expressed in NSCLC and (2) they were significantly anti-

correlated (Spearman ρ < −0.3, *P*-value < 0.05, and false discovery rate [FDR] < 0.1) (see Materials and Methods). Based on the available miRNAs in TargetScan database and the pairs that satisfied our filtering criteria above, we obtained 744 miRNA–target gene pairs from 19 up-regulated miRNAs and 301 down-regulated genes in NSCLC compared with the normal lung tissue. In addition, we obtained 575 miRNA–target gene pairs from 17 down-regulated miRNAs and 221 up-regulated genes in NSCLC compared with the normal lung tissue (Fig. 1A). We employed the same strict criteria for identifying miRNA-mediated TF repression. This resulted in 89 miRNA–TF pairs from 29 miRNAs and 31 TFs (see Materials and Methods). To find the regulation of an active

TABLE 1. Summary of the integrated regulatory relationship among miRNAs, TFs, and genes

Relationship	Potential regulation type	# pairs	# miRNAs	# genes	# TFs	Detection method[a]
miRNA–gene	Repression	1319	36	522	—	A
miRNA–TF	Repression	89	29	—	31	B
TF–gene	Activation	799	—	293	48	C
TF–gene	Repression	576	—	252	46	C
TF–miRNA	Activation	39	16	—	23	C
TF–miRNA	Repression	46	19	—	20	C

[a]A: Differential expression, TargetScan, and Spearman rank correlation. B: Differential expression, TargetScan, Spearman rank correlation, and TRANSFAC. C: Differential expression, Spearman rank correlation, TRANSFAC, and MATCH.

TF to genes or miRNAs, we explored the active TFs and their binding profiles from the TRANSFAC Professional database and predicted TF-binding sites using its Match software (Kel et al. 2003). To reduce the false positive predictions, we required the predicted pairs to be conserved among humans, mice, and rats (Guo et al. 2010) (see Materials and Methods). Again, to assign the direction of the regulation, we combined the TF–target prediction results and *Discovery_data*. The criteria for TF-mediated repression of gene/miRNA were the same as those of the miRNA-mediated gene/TF repression procedure, which was described above. For the TF-mediated gene/miRNA activation, we required that the TF and the predicted target need to be concordantly expressed (either up- or down-regulated) in NSCLC and that they must be positively correlated (Spearman $\rho > 0.3$, P-value < 0.05, and FDR < 0.1) (see Materials and Methods). We predicted 576 TF–target gene pairs and 46 TF–target miRNA pairs that were negatively correlated and 799 TF–target gene pairs and 39 TF–target miRNA pairs that were positively correlated (see Table 1).

miRNA–TF co-regulatory network in NSCLC with 57% validated regulations

Our network was constructed based on the transcriptional regulation of TFs tightly coupled with the post-transcriptional regulation of miRNAs. These two types of gene regulators frequently form three-node FFLs (i.e., each FFL has a TF, a miRNA, and a common target gene). To study the DE miRNA and active TF-mediated co-regulation in NSCLC, we utilized their three-node FFL motifs. We then performed the randomization test using the tool FANMOD (Wernicke 2006; Wernicke and Rasche 2006) to assess whether a three-node FFL motif was unlikely generated from a network by chance (Supplemental Text S1). Consequently, we identified four types of overrepresented FFLs in NSCLC (P-value $< 10^{-2}$, randomization test); they were named motifs A, B, C, and D (Table 2). There were a total of 417 FFLs belonging to these four types of FFL motifs (Table 2; Fig. 1E).

In motif A, the TF induces target gene expression at the transcriptional level, and the miRNA represses both TF and

gene expression at the post-transcriptional level. Motifs B and C share a common design structure: The TF regulates the miRNA and the protein-coding gene at the transcriptional level, and the miRNA represses the target gene at the post-transcriptional level. However, in motif B, the TF induces miRNA expression and represses gene expression, while in motif C, the TF represses miRNA expression and induces gene expression. In motif D, the regulators mutually repress each other; the TF induces target gene expression at the transcriptional level, and the miRNA represses target gene expression at the post-transcriptional level. We constructed a miRNA–TF co-regulatory network by merging these 417 FFLs, as mentioned above. The network consisted of 32 DE miRNAs, 31 active TFs, 147 DE genes, and 688 edges. Expression patterns, fold-changes, and adjusted P-values for these miRNAs, TFs, and genes are provided in Supplemental Tables S6–S9. The detailed information about the direction of biological regulation in terms of expression correlation for the 688 edges is provided in Supplemental Table S10.

Among the 688 edges, 578 were commonly detected in the three independent validation data sets. A Venn diagram was provided in Supplemental Figure S2. To define if one edge was validated, we required that the direction of expression correlation of that edge should be consistent in at least two validation data sets. Furthermore, the summary of the three correlation P-values, combined by the Fisher's method, should be < 0.05 with its FDR < 0.1. The detailed network edge validation strategy is described in Materials and Methods. Among the 578 edges, 332 (57.44%) edges were validated by independent validation data sets. We further divided the validated edges into three groups according to the number of validated data sets and related significance. The first and second groups included 36 (6.23%) and 133 (23.01%) edges that were strongly validated by three and two validation data sets, respectively. The last group included 163 edges (28.20%) that were validated by two validation data sets. Among these 163 edges, 159 were strongly validated by one independent validation data set. We defined a validated edge as "strongly validated" if it showed the expression correlation P-value of < 0.05. Detailed validation results are provided in Supplemental Figure S3 and Supplemental Table S11. Throughout the manuscript, we synonymously used the terms "validated edges" and "validated targets" to represent a reproducible biological regulation. For clarity, we assigned different colors and line styles to explain different groups of edges.

Systematic filtering of critical miRNAs in NSCLC

A general view of the regulatory network can be observed by calculating the node degree (connectivity) and its

TABLE 2. Summary of significantly enriched three-node FFLs based on DE miRNAs, DE genes, and active TFs

Motif	#FFLs	Number of nodes			Number of links			
		TFs	miRNAs	Genes	miRNA-Gene	miRNA-TF	TF-Gene	TF-miRNA
A	223	23	26	101	180	63	146	0
B	80	18	13	57	64	0	76	27
C	70	13	17	48	57	0	63	25
D	44	9	8	23	34	13	34	13
Total	417	31	32	147	273	76	274	65

For each of the hub miRNAs and TFs, we conducted a functional analysis of its targets using the software WebGestalt (Zhang et al. 2005). We found a profiling set of 12 hub miRNAs and six hub TFs whose targets were significantly enriched with 12 gene ontology (GO) terms that are closely related to cancer-specific biological processes (hypergeometric test P-value < 0.05, followed by the Benjamini-Hochberg [BH] method [Benjamini and Hochberg 1995] for multiple test correction) (Fig. 2A). Functional enrichment of target genes and TFs in a number of cancer-related biological processes suggest that these hub regulators might have important roles in cancer development. Figure 2A shows that targets of three up-regulated miRNAs in NSCLC (miR-9-5p, miR-183-5p, and miR-200b-3p) were enriched in the apoptotic process or programmed cell death. The down-regulation of their targets in NSCLC suggests that the apoptotic process might be hampered because of the involvement of these miRNAs. In contrast, targets of miR-30a-5p were significantly enriched in the negative regulation of the apoptotic process. Targets of miR-30a-5p were up-regulated in NSCLC, which might be due to the down-regulation of miR-30a-5p, and the consequence of this would be reduced apoptosis. Furthermore, a large set of eight miRNAs (miR-130b-3p, miR-17-5p, miR-200c-3p, miR-183-5p, miR-200b-3p, miR-96-5p, miR-9-5p, and miR-182-5p) and three TFs (E2F3, ZEB1, and EGR2) were significantly associated with at least one of the

distribution, which are basic topological network measures (Barabasi and Oltvai 2004). The degree values for TFs, miRNAs, and genes in this miRNA–TF co-regulatory network ranged from 2 to 33, 2 to 29, and 2 to 14, respectively. Degree distributions of all nodes were right-skewed, indicating that a high degree could be observed for a small portion of nodes, and these nodes might be considered as hubs in the network (Supplemental Fig. S4). Based on the definition proposed by Yu et al. (2004), we pinpointed hub nodes in the network using the degree cutoff value of 10. Thus, we obtained 40 hub nodes, including 20 miRNAs, 17 TFs, and three protein-coding genes.

FIGURE 2. Critical hub regulators that are potentially involved in cancer progression. (*A*) Enriched gene ontology (GO) terms in the targets of 18 hubs (six TFs and 12 miRNAs). (*B*) Scatter plot of hub nodes illustrating the percentage of validated targets and cancer-related GO terms. The *x*-axis represents the number of cancer-related GO terms associated with the targets for each miRNA or TF. The *y*-axis represents the percentage of validated targets observed for each node. Squares and triangles represent TFs and miRNAs, respectively. Red corresponds to up-regulation, and green corresponds to down-regulation.

following biological processes: cell migration, cell differentiation, and intracellular signal transduction. These biological processes are associated with epithelial-to-mesenchymal transition (EMT), a process thought to be essential in cancer metastasis (Kalluri and Weinberg 2009; Liu 2010). The numbers of targets for all miRNAs presented in this regulatory network are provided in Supplemental Figure S5.

Remarkably, among the 12 hub miRNAs and six hub TFs, eight miRNAs (miR-130b-3p, miR-17-5p, miR-200c-3p, miR-183-5p, miR-200b-3p, miR-96-5p, miR-9-5p, and miR-182-5p) and four TFs (ZEB1, EGR2, HMGA1, and SMAD3) had at least 50% of their targets validated by independent validation data sets (Fig. 2B). The proportions of validated targets of three miRNAs (miR-130b-3p, miR-200c-3p, and miR-17-5p) were >90%. The proportions of other two miRNAs (miR-183-5p and miR-200b-3p) and the TF ZEB1 were >80%. The higher proportion of validated targets for a miRNA or a TF may increase the possibility of having true positive associations with cancer-related biological functions. Conversely, two miRNAs (miR-30d-5p and miR-30b-5p) and one TF (SMAD1) had a considerably lower proportion of validated targets (<30%) (Fig. 2B) and might be susceptible to false-positive predictions.

Next, we pinpointed to uncover critical biological regulations that might govern the tumorigenesis of NSCLC. Recently, pathway analysis has been reported as a useful approach to assess the biological mechanisms involved in the disease pathogenesis (Wang et al. 2011). We examined enriched pathways in this regulatory network. For the 31 TFs and 147 genes, we identified 32 significantly enriched pathways (adjusted P-value $< 10^{-2}$, hypergeometric test followed by BH multiple testing correction) from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto 2000). A list of these pathways is provided in Supplemental Table S12. Among these 32 pathways, 21 (65.62%) are related to cancer. Of these cancer-related pathways, six are known to have association with NSCLC progression, including Wnt, NSCLC, cell cycle, TGF-β, ErbB, and MAPK.

Potential suppression of TGF-β signaling by critical hub miRNAs

TGF-β signaling plays a major role in tumorigenesis (Pino et al. 2010; Yang and Yang 2010). Based on the canonical KEGG pathway database (Kanehisa and Goto 2000), our miRNA–TF co-regula-

tory network could recruit five genes in the TGF-β signaling pathway: four SMAD family TFs (SMAD1, SMAD2, SMAD3, and SMAD7) and the gene *TGFBR2*. We extracted a sub-network from the miRNA–TF co-regulatory network by merging all the FFLs that included at least one of these TFs or *TGFBR2*. This sub-network included 194 edges, 22 miRNAs, six TFs, and 36 genes (Fig. 3A). The sub-network showed some meaningful results. For example, the two hub TFs Smad2/3 that appeared in this sub-network are well-known surrogates for TGF-β activation (Nyati et al. 2011). Furthermore, down-regulation or loss of the tumor suppressor gene *TGFBR2* in cancer impedes TGF-β-mediated tumor suppressor functions (Jeon and Jen 2010).

We used the software CFinder (Palla et al. 2005) to identify tightly connected overlapping interaction modules in NSCLC. This analysis resulted in four modules (Modules 1–4) (Fig. 3A) whose size was more than five nodes. Modules 1, 2, 3, and 4 consisted of 12.5%, 33.04%, 50.89%, and 88.89% validated targets, respectively (Fig. 3B). Module 4 appeared to have the most confident regulatory interactions in which a significantly enriched number of edges were validated by independent data sets (P-value $= 5.3 \times 10^{-3}$, randomization test, see Materials and Methods). It consisted of the
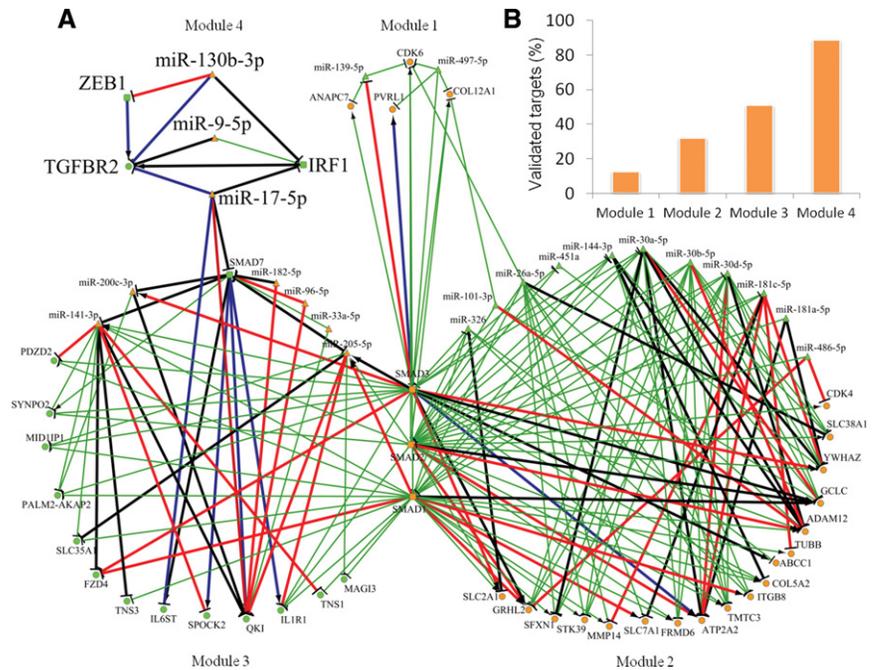
FIGURE 3. miRNA- and TF-mediated sub-network specific to the TGF-β signaling pathway. (*A*) The network consisted of four modules (Modules 1–4). Module 4 contains three miRNAs (miR-17-5p, miR-9-5p, and miR-130b-3p), two TFs (IRF1 and ZEB1), and one gene (*TGFBR2*). Edges with different colors represent their validation status as assessed using three independent data sets. Blue edge: Regulations were strongly validated by three independent data sets; black edge: strongly validated by two independent data sets; red edge: validated by two independent data sets; and, green edge: either not validated or do not exist commonly in all the three validation data sets. Strong validation is defined as a Spearman rank correlation P-value < 0.05. (*B*) Bar plot illustrates percentage of validated targets for four modules. The x-axis represents the four modules. The y-axis represents percentage of validated targets observed for each module.

core TGF-β pathway gene *TGFBR2* and five regulators (three miRNAs: miR-9-5p, miR-130b-3p, and miR-17-5p; two TFs: ZEB1 and IRF1). Among these five regulators, we identified four (miR-9-5p, miR-130b-3p, and miR-17-5p, and ZEB1) as cancer-causing critical hub nodes. The direction of the network edges demonstrated that *TGFBR2* could be positively regulated by the two TFs at the transcriptional level and repressed by the three miRNAs at the post-transcriptional level (Fig. 3A). Of note, out of the five edges connected to *TGFBR2*, three edges were strongly validated by all three independent validation data sets (miR-17-5p-*TGFBR2*, miR-130b-3p-*TGFBR2*, and ZEB1-*TGFBR2*), and two were strongly validated by two independent validation data sets (miR-9-5p-*TGFBR2* and IRF1-*TGFBR2*). Repression of *TGFBR2* expression could occur through up-regulation of its targeting miRNAs.

A previous study has confirmed the up-regulation of miR-17-5p in NSCLC compared with the normal lung tissue (Navarro et al. 2009). Moreover, *TGFBR2* was reported as a direct target of miR-17-5p (Dews et al. 2010). So far, the expression pattern for miR-9-5p in lung tumor versus non-tumor samples has been inconsistently reported: up-regulation (Crawford et al. 2009; Kang et al. 2013; Xu et al. 2014), down-regulation (Yanaihara et al. 2006), or insignificant change in expression (Jusufovic et al. 2012). To the best of our knowledge, there is no report on miR-130b-3p expression in NSCLC.

Tumor suppressor gene *TGFBR2* regulated by miR-9-5p and miR-130b-3p cooperatively

To validate the regulation of *TGFBR2* in Module 4, we first performed qRT-PCR to assess the expression of miR-9-5p, miR-130b-3p, and *TGFBR2* in 14 human NSCLC tumor and 10 normal lung tissue samples. Relative expression of miR-9-5p (Wilcoxon rank-sum test, $P = 5.76 \times 10^{-4}$) and miR-130b-3p (Wilcoxon rank-sum test, $P = 3.35 \times 10^{-4}$) was significantly higher in NSCLC when compared with normal lung tissue (Fig. 4). In contrast, the expression of *TGFBR2* mRNA was significantly lower (Wilcoxon rank-sum test, $P = 3.569 \times 10^{-6}$) in NSCLC compared with the

normal lung tissue (Fig. 4). Additionally, we observed significant negative correlations between the expression levels of miR-9-5p and *TGFBR2* (Spearman $\rho = -0.47$, *P*-value = 0.02), as well as between miR-130b-3p and *TGFBR2* (Spearman $\rho = -0.54$, *P*-value = 7.13×10^{-3}).

We then tested whether *TGFBR2* is a direct target of miR-9-5p and miR-130b-3p. A luciferase expression plasmid with the 3′ UTR of *TGFBR2* was co-transfected with mimics of miR-9-5p, miR-130b-3p, or miR-10b-5p, or control RNA, in NIH3T3 cells. The cells with miR-9-5p mimic had a significant reduction in luciferase activity, compared with the RNA control transfected cells (*P*-value = 0.01) (Fig. 5). Cells with miR-130b-3p mimic also showed a significant reduction in luciferase activity compared with the control (*P*-value = 0.026) (Fig. 5). miR-10b-5p is not predicted to bind to the 3′ UTR of *TGFBR2*. In our experiments, cells with miR-10b-5p mimic did not show a reduction in luciferase activity, indicating that the reductions in luciferase activity we detected with miR-9-5p and miR-130b-3p were specific. These data indicate that *TGFBR2* is likely a direct target of miR-9-5p and miR-130b-3p. Interestingly, when miR-9-5p and miR-130b-3p were transfected together, they reduced luciferase activity more than either one alone (*P*-value = 2.0×10^{-6}) (Fig. 5). This result suggested that these two miRNAs might have functioned cooperatively to suppress *TGFBR2* expression.

DISCUSSION

In this study, we constructed a miRNA and TF co-regulatory network, which provided several important insights into the etiology of NSCLC. Our framework started with the identification of significantly DE genes and miRNAs in NSCLC, since differential miRNA and gene expression have functional relevance in tumorigenesis. Furthermore, we restricted the TFs to those that are potentially active in NSCLC. We inferred regulatory relationships, including the direction of biological regulations among the DE miRNAs, active TFs, and DE genes using a panel of computational tools, and matched miRNA and gene expression profiles from carefully selected *Discovery_data*.
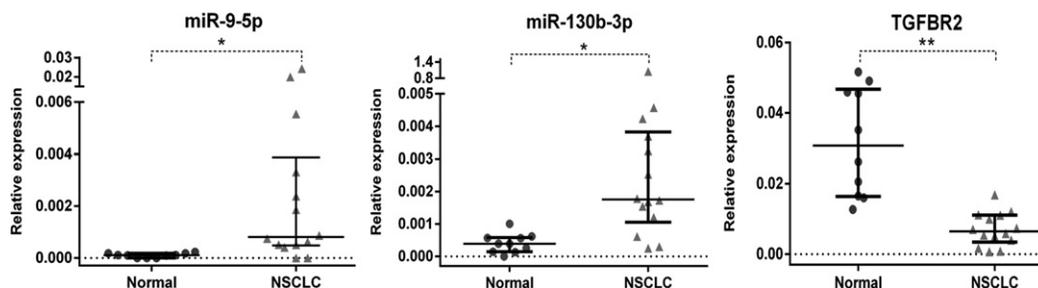
FIGURE 4. Increased expression of miR-9-5p and miR-130b-3p and decreased expression of *TGFBR2* in NSCLC. Expression levels of miR-9-5p, miR-130b-3p, and *TGFBR2* were assessed by qRT-PCR in 14 NSCLC stage I tumor samples and 10 normal lung samples. Data are relative to an internal RNA control and represent median with interquartile range. A statistical analysis was performed using the Wilcoxon rank-sum test. Statistically significant differences were marked as (*) *P*-value < 10^{-3} or (**) *P*-value < 10^{-5}.
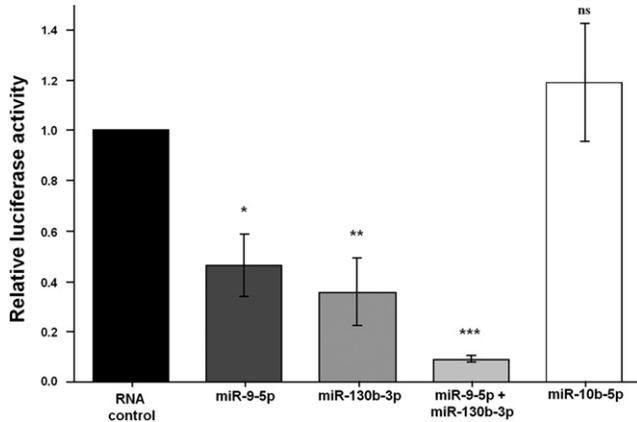
FIGURE 5. *TGFBR2* is a direct target of miR-9-5p and miR-130b-3p. Luciferase activity was measured in NIH3T3 cells transfected with the indicated miRNA mimic or RNA control, a plasmid with *TGFBR2* 3′ UTR, and a β-galactosidase plasmid. Luciferase activity was relative to β-galactosidase activity, which controlled for transfection efficiency for each. ±SEM; $n = 4$; (*) P-value = 0.01, (**) P-value = 0.026, (***) P-value = 2.0×10^{-6}; ns, not significant; t-tests. miR-10b was not predicted to bind to the 3′ UTR of *TGFBR2*, and thus, served as an additional negative control.

One main issue in such a large-scale computational data analysis is to control the false positives during the construction of regulatory relationships. To minimize the effects of false positives, we applied the following procedures. First, we used the most annotated databases and appropriate software and algorithms to conduct the prediction. Second, we applied stringent parameters in the prediction of miRNA to gene/TF and TF to gene/miRNA relationships. Third, we required the observed regulatory relationships to be conserved across multiple mammalian genomes. Fourth, and more importantly, we imparted the direction of the biological regulation in order to pinpoint only biologically meaningful regulatory relationships. Our framework could potentially detect meaningful regulatory relationships and might be applied to other complex diseases for the purpose of deciphering their regulatory systems and identifying critical miRNAs. Due to the rapid growth of high-throughput expression profiling studies, we could access multiple miRNA and gene expression data sets involving NSCLC samples and even their matched normal samples (see Materials and Methods; Supplemental Table S1). We addressed the reproducible biological regulations in this miRNA–TF co-regulatory network through an application of a novel network edge validation strategy using three independent validation data sets. These reproducible regulatory relationships are expected to more likely reflect the true biological regulations in a cellular system (Langfelder et al. 2011; Dutta et al. 2012).

One important outcome of this comprehensive study is the eight hub miRNAs (miR-130b-3p, miR-17-5p, miR-200c-3p, miR-200b-3p, miR-183-5p, miR-182-5p, miR-96-5p, and miR-9-5p) that were identified with a higher proportion of targets as validated by independent validation data sets

(in the range of 50%–94.5%). Target genes/TFs for these eight miRNAs were found to be significantly associated with multiple cancer-related biological processes. These eight miRNAs were significantly up-regulated in NSCLC compared with normal samples. We conducted qRT-PCR analyses and confirmed the up-regulation of miR-130b-3p and miR-9-5p in NSCLC (Fig. 4). Up-regulation for the other six miRNAs (miR-17-5p, miR-200c-3p, miR-200b-3p, miR-183-5p, miR-182-5p, and miR-96-5p) had been established by qRT-PCR analysis in previous studies (Supplemental Table S13). Collectively, these results demonstrate that our regulatory network analysis is effective to identify oncogenic miRNAs.

An in-depth literature search elucidated that there are 16 (~9% of total genes and TFs) candidate tumor suppressor genes/TFs in this regulatory network. Intriguingly, the analysis of RNA-Seq expression profiles from *Discovery_data* could explain that these 16 tumor suppressors were significantly down-regulated (at least twofold change with BH adjusted P-value < 10^{-4}) in NSCLC compared with the normal lung tissue (Supplemental Table S14). Down-regulation of tumor suppressors in cancer tissues or cell lines might support the general understanding that they are involved in tumorigenesis. Interestingly, these 16 tumor suppressor nodes in our regulatory network connected with the eight hub miRNAs, mentioned above, and formed a sub-network that included 32 edges (Fig. 6A). Among these 32 edges, 30 (93.7%) were validated using independent validation data sets. We performed an empirical re-sampling approach to assess whether the observed edge validation result is by chance. The empirical P-value was 2.0×10^{-4}, suggesting that this is unlikely by randomness. miR-17-5p targeted eight tumor suppressor genes/TFs (*TGFBR2*, *RASSF*2, *FAT*4, *SPRY*4, *DAB*2, *QKI*, IRF1, and EGR2). Among these eight genes/TFs, *TGFBR2*, *RASSF*2, IRF1, and EGR2 are experimentally validated targets, as reported previously (see Supplemental Table S15). miR-9-5p targeted six tumor suppressor genes/TFs (*TGFBR2*, *EFEMP*1, *LIFR*, *RPS6KA*2, *STARD*13, and IRF1). Recently, Namlos et al. (2012) reported an anti-correlation between miR-9-5p and *TGFBR2* in osteosarcoma. The remaining six hub miRNAs were potentially involved in the repression of a range of two to five tumor suppressor genes/TFs. Moreover, from the *Discovery_data* we observed that the tumor suppressor genes/TFs and the regulating hub miRNAs have a stronger anti-correlation compared with the remaining miRNA–target pairs in the network (P-value = 6.95×10^{-3}, Wilcoxon rank-sum test) (Fig. 6B). Collectively, these lines of evidence further supported that these eight miRNAs might have oncogenic potential in the pathogenesis of NSCLC.

Among these eight hub miRNAs, our follow-up experimental results confirmed that *TGFBR2* is a direct target of both miR-9-5p and miR-130b-3p. Intriguingly, we determined that miR-9-5p and miR-130b-3p cooperatively suppressed *TGFBR2* 3′ UTR reporter activity (Fig. 5), indicating that these two miRNAs could act together to effectively block *TGFBR2* expression. The TGF-β signaling pathway is
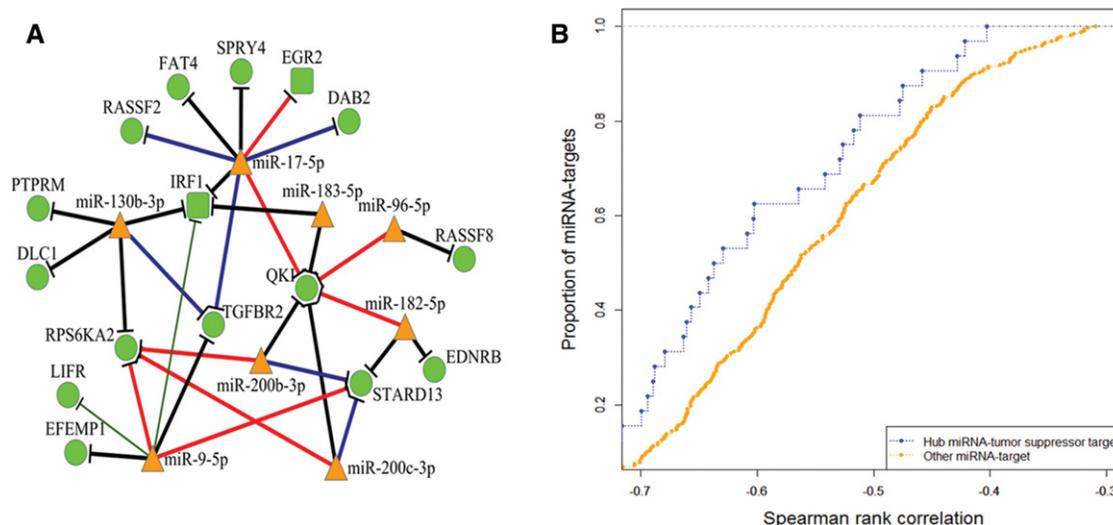
FIGURE 6. Association between hub miRNAs and candidate tumor suppressors in NSCLC. (*A*) Combinatorial interactions between eight hub miRNAs and 16 tumor suppressor genes/TFs. (*B*) Empirical cumulative distribution plot of Spearman rank correlation. Significantly higher anti-correlations were observed in the pairs of hub miRNAs and tumor suppressors compared with the remaining miRNA–target pairs (Wilcoxon rank-sum test, *P*-value = 6.95×10^{-3}) within the miRNA–TF co-regulatory network.

reportedly impeded by the loss of *TGFBR2* (Anumanthan et al. 2005; Munoz et al. 2006; Biswas et al. 2008; Jeon and Jen 2010). Thus, our results provided additional evidence of alterations in the TGF-β signaling pathway in NSCLC through miRNA co-regulation.

There are some limitations in this study. First, we employed strict criteria to identify regulatory relationships among miRNAs, TFs, and genes. While this practice is common to reduce false positives, it might result in missing some true positive regulations. Second, the study was conducted based on the available patient samples from both lung squamous cell carcinoma and lung adenocarcinoma and elucidated critical regulatory mechanisms for NSCLC. While we pointed this out as a limitation, several studies regarded these two subtypes together in order to infer critical biomolecules that have potential implications for the pathology of NSCLC (Raponi et al. 2006; Ma et al. 2011; Xu et al. 2014). For example, after analyzing 116 tumor samples (69 squamous cell carcinoma and 47 adenocarcinoma) and adjacent normal lung tissue samples, Xu et al. (2014) demonstrated that high miR-9-5p expression is associated with poorer progression-free survival (*P*-value < 10^{-3}) and overall survival (*P*-value < 10^{-3}) in the NSCLC patients. Third, in this study, the direction of biological regulation was based on the available tumor and normal samples from *Discovery_data*. However, complex diseases incorporate many changes at the genomic and transcriptomic levels through a gradual process (Zheng et al. 2011). Matched miRNA and gene expression profiles from time point data could be used to address meaningful node association changes over different time points. However, a lack of such comprehensive data sets is a current limitation to decipher the accurate regulatory mechanism for most of the complex diseases.

There is one scope to extend the current computational framework by integrating other regulation information. We may include another filter step that regards gene expression as a function of DNA methylation, copy number variation, miRNA, and TF expression. Using multivariate regression model we could evaluate miRNA–gene or TF–gene expression associations in the presence of DNA copy number and promoter methylation aberrations that extensively influence gene expression.

In summary, our miRNA–TF co-regulatory network analyses elucidated some critical regulators and clues for the regulatory mechanisms of NSCLC. We confirmed up-regulation of miR-9-5p and miR-130b-3p and down-regulation of tumor suppressor gene *TGFBR2* in NSCLC. We experimentally validated miR-9-5p and miR-130b-3p mediated co-operative regulation of *TGFBR2*. These experimental results, along with previous studies, support that miR-9-5p and miR-130b-3p might play critical roles in the pathology of NSCLC. Consequently, our study provided a new strategy that combined bioinformatics analyses of miRNA–TF co-regulation with experimental validation to study complex diseases.

MATERIALS AND METHODS

Expression profiles for genes and miRNAs

We collected four data sets, including one for discovery purpose and three for validation purposes.

Discovery_data

The Cancer Genome Atlas (TCGA) utilized RNA sequencing (RNA-Seq) to quantify gene expression for lung squamous cell carcinoma, a subtype of NSCLC (Hammerman et al. 2012). To explain the

abundance of gene expression, TCGA provided two measures—the read count and tau value. We transformed the tau value to transcripts per million (TPM) as recommended by Li and Dewey (2011). TCGA also archived the NSCLC miRNA-Seq data to quantify miRNA expression. The data consisted of read counts and reads per million (RPM) values for mature miRNAs. We utilized read counts to perform the differential expression analysis. We utilized RPM and TPM to compute the expression correlation between a regulator and a target molecule.

In February 2013, TCGA archived 16 NSCLC patient samples and 16 matched normal samples for which the gene expression and miRNA expression profiles were available. Specifically, both the RNA-Seq and the miRNA-Seq expression profiles from the same samples were retrieved from TCGA (https://tcga-data.nci.nih.gov/tcga/) (see Supplemental Tables S1, S2 for more details).

Validation_data1

Matched miRNA and gene expression profiles for a set of 57 lung squamous cell carcinoma samples were obtained from the Gene Expression Omnibus (Edgar et al. 2002). Processed and quantile normalized miRNA expression profiles were available in the GEO database (accession number: GSE16025), submitted by Raponi et al. (2009). Microarray gene expression profiles (GEO accession number: GSE4573) were normalized according to the instructions provided by Raponi et al. (2006) (see Supplemental Tables S1, S3 for more details).

Validation_data2

An integrative lung cancer study was carried out by Nymark et al. (2011) using both miRNA and matched gene expression profiling data. The study considered lung cancer patients that were "highly exposed" and "nonexposed" to asbestos. miRNA and matched microarray gene expression profiles were provided by the authors (P Nymark and S Knuutila, pers. comm.). Both the data sets were processed and normalized accordingly as described by Nymark et al. (2011). We considered seven asbestos-nonexposed NSCLC samples and corresponding matched normal samples that had both miRNA and gene expression profiling information (see Supplemental Tables S1, S4 for more details).

Validation_data3

From the original publication of TCGA (Hammerman et al. 2012), matched miRNA and gene expression profiles from 154 lung squamous cell carcinoma patient samples were retrieved. This data set is completely independent from the *Discovery_data* and was not used to construct the regulatory network. Normalized miRNA and matched gene expression profiles were directly obtained from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/) (see Supplemental Tables S1, S5 for more detail).

Differential expression analysis using *Discovery_data*

Both miRNA-Seq and RNA-Seq expression profiles from the *Discovery_data* were analyzed for identifying DE miRNAs and genes through the R/Bioconductor package edgeR (Robinson et al. 2010), which was designed to analyze digital miRNA and gene expression data. The read counts were imported into *EdgeR* for differential expression analysis. The data were normalized based on negative bino-

mial distribution. The differential expression of miRNAs/genes between tumor and normal samples was assessed by estimating an exact test *P*-value, which is similar to the Fisher's exact test. Then, the results were further adjusted using the BH multiple testing correction method (Benjamini and Hochberg 1995). Two types of differential expression, namely up-regulation and down-regulation, were defined according to the following criteria: twofold change and adjusted P-value $< 10^{-4}$.

miRNA-mediated gene/TF repression

We integrated TargetScan (release 6.2) prediction results with the miRNA and gene expression profiling data set (*Discovery_data*) to identify miRNA-mediated repression of target genes/TFs. TargetScan searches for the presence of conserved and nonconserved miRNA–target sites in the 3′ UTR of the target transcript (Friedman et al. 2009). The algorithm distinguishes conserved and poorly conserved targets of each miRNA by measuring the "probability of conserved targeting" (P_{CT}) values of the target sites. We restricted our analysis to only conserved miRNA–target pairs using "Predicted Conserved Targets Info" file, which was provided by the TargetScan database. We placed miRNAs and their predicted target genes into group 1 in which miRNAs were down-regulated and genes were up-regulated. We placed miRNAs and their predicted target genes into group 2 in which miRNAs were up-regulated and genes were down-regulated (see Fig. 1A). Furthermore, we selected miRNA–gene pairs from the two groups that were anti-correlated, had a Spearman rank correlation of $\rho < -0.3$, and had a correlation *P*-value of <0.05. Finally, we performed multiple testing correction using the BH method (Benjamini and Hochberg 1995) and required the adjusted *P*-values (FDR) <0.1 to define significant anti-correlations.

To obtain miRNA-mediated TF repression, we utilized all the TFs that were potentially active (see description below) in NSCLC and retrieved predicted miRNA–TF pairs using TargetScan (Friedman et al. 2009). We placed down-regulated miRNAs and up-regulated predicted target TFs into one group and up-regulated miRNAs and down-regulated predicted target TFs into another group. Finally, we extracted miRNA–TF pairs from the two groups that were significantly anti-correlated (Spearman rank correlation of $\rho < -0.3$, *P*-value < 0.05, and FDR < 0.1).

TF-mediated gene/miRNA regulation

We extracted the promoter region (−1500/+500 around TSS) of human genes and miRNAs from the UCSC Table Browser (Kent et al. 2002) as explained by Sun et al. (2012). We performed binding site searches using the software Match (Kel et al. 2003), which was available in the TRANSFAC professional version (release 2011.4) (Matys et al. 2006). We utilized pre-calculated cutoffs and a high-quality matrix to minimize false positive matches. To restrict the search, we selected a TF by requiring a core score of 1.00 and a matrix score of 0.95. To further reduce false positive predictions, we required the predicted pairs to be conserved among human, mouse, and rat data (Guo et al. 2010).

Active TF identification

To define a TF as active or not, we examined the expression of the TF-encoding gene and its targets in NSCLC. If the TF-encoding gene was DE (category 1) or most of its targets were DE (category

2), the TF was considered to be active (Sohler and Zimmer 2005; Essaghir et al. 2010; Liu et al. 2010; Naeem et al. 2011). To identify the active TFs by their targets' DE status, we applied the hypergeometric test to estimate the significance of the TFs based on the observed overrepresentation of their targets among the DE genes or DE miRNAs. We obtained 479 DE targets (450 DE genes and 29 DE miRNAs) of TFs. We also obtained 2495 non-DE targets (less than twofold change and adjusted P-value > 0.25) of TFs. DE and non-DE status were determined from the *Discovery_data*. The P-value was computed according to the following hypergeometric formula:

$$p = \sum_{i=k}^{i=n} \binom{m}{i}\binom{N-m}{n-i}/\binom{N}{n},$$

where m is the number of DE miRNAs and genes, n is the number of miRNAs and genes targeted by a TF of interest, N is the total number of DE and non-DE miRNAs and genes, and k is the number of DE miRNAs and genes targeted by the TF of interest. We obtained 19 active TFs in category 2 whose significant number of targets (hypergeometric test, P-value < 0.05) were DE. Lists of active TFs are provided in Supplemental Tables S7, S8.

Active TF-mediated gene/miRNA regulation

We selected active TFs and their predicted target genes if both the TFs and their target genes were either up- or down-regulated in NSCLC. For all such TF and target gene pairs, we computed their Spearman rank correlations. If the pairs were significantly positively correlated (Spearman $\rho > 0.3$, P-value < 0.05, and FDR < 0.1), we selected them as TF-mediated activations of target genes. However, pairs were selected as TF-mediated repression of target genes if they had both of the following traits: (1) Active TFs were up-regulated in NSCLC and their predicted target genes were down-regulated in NSCLC (or vice-versa), and (2) they were significantly anti-correlated (Spearman $\rho < -0.3$, P-value < 0.05, and FDR < 0.1). Here, we applied FDR threshold value 0.1 to define significantly positive or negative correlations.

We selected the active TFs and their predicted target miRNAs if both the TFs and the miRNAs were either up- or down-regulated in NSCLC. If the pairs were significantly positively correlated (Spearman $\rho > 0.3$, P-value < 0.05, and FDR < 0.1), we selected these pairs as TF-mediated activations of target miRNAs. On the other hand, if the TFs were up-regulated and the predicted target miRNAs were down-regulated in NSCLC (or vice-versa), and the pairs were significantly negatively correlated (Spearman $\rho < -0.3$, P-value < 0.05, and FDR < 0.1), the pairs were selected as TF-mediated repressions of target miRNAs.

Network edge validation strategy using three independent validation data sets

Using the *Discovery_data*, we computed a Spearman rank correlation to impart the direction of regulation for each regulator–target pair. Next, we used three validation data sets to assess whether the expression correlation for a regulator–target pair is consistent in independent studies. Therefore, for one regulator–target pair, we obtained three Spearman rank correlation scores and corresponding P-values from the three validation data sets. We employed Fisher's method to combine the three P-values and obtained a summary

P-value. Furthermore, we performed multiple testing corrections using the FDR method (Benjamini and Hochberg 1995). To define if an edge was validated, we required that the direction of the regulation (positive/negative correlation score observed in *Discovery_data*) should be consistent in at least two independent validation data sets; and, the combined P-value measured by Fisher's combined method should be <0.05 with its FDR < 0.1. Detailed validation results are shown in Supplemental Figure S3. To perform the Fisher's combined method, we used an R/Bioconductor package "survcomp" (Schroder et al. 2011).

Network randomization analyses

We applied an empirical re-sampling approach in order to evaluate the enrichment of the validated edges in a sub-network. Assuming that among n edges in the sub-network, m edges were validated by the independent validation data sets. From the miRNA–TF co-regulatory network, we took into account 578 edges that were common in the three validation data sets. We first excluded the observed sub-network from the 578 edges and then randomly selected n number of edges. We compared the occurrences of number of validated edges in the randomly generated sub-network and repeated the process 10,000 times. We counted the number of randomly generated sub-networks (N) whose number of validated edges was greater than m. Finally, we calculated an empirical P-value by $N/10{,}000$; here, the P-value indicates whether the observed sub-network with m-validated edges was unlikely generated by chance.

Quantitative real-time PCR with human samples

We obtained de-identified frozen human samples (14 lung adenocarcinoma and 10 normal lung) from the Lung Biorepository at Vanderbilt University Medical Center. All samples were collected prior to chemotherapy and/or radiotherapy, and all tumors were in stage I NSCLC. Total RNA was isolated using Trizol (Life Technologies) following the manufacturer's protocol. miRNA analysis was performed as previously reported by Arrate et al. (2010). Briefly, the Taqman miRNA reverse transcription kit (Life Technologies) was used to make miRNA-specific cDNA. The expression of miR-9-5p, miR-130b-3p, and RNU48 (as an endogenous control) was measured in triplicate on an Applied Biosystems instrument with Taqman primer probes (Life Technologies) specific for each. The expression of *TGFBR2* and β-*ACTIN* mRNA was measured in triplicate as previously described by Wang et al. (2008) using the Super-Script III Reverse Transcriptase kit (Life technologies) and SYBR Green (Qiagen). The *TGFBR2* primers (forward: 5′-GTAGCTCT GATGAGTGCAATGAC-3′; reverse: 5′-CAGATATGGCAACTCCC AGTG-3′) were from the Harvard Primer Bank, and the β-*ACTIN* primers (forward: 5′-GGATGCAGAAGGAGATCA-3′; reverse: 5′-CTAGAAGCATTTGCGGTG-3′) were published previously (Gonzalez et al. 2010). We calculated the expressions of miR-130b-3p, miR-9-5p, and *TGFBR2* relative to endogenous controls, which were presented as $2^{-\Delta Ct}$.

Luciferase assay

Luciferase assays were conducted as previously reported by McGirt et al. (2014), except in the following cases. NIH3T3 cells were co-transfected with 100 ng of psiCHECK-2-*TGFBR2* 3′ UTR plasmid

(31883, Addgene) (Subramanyam et al. 2011), 200 nM miRNA mimic (miR-9-p, miR-130b-3p, or miR-10b-5p) or control RNA, and 100 ng pMIR-REPORT β-galactosidase for normalizing transfection efficiency. Luciferase activity and β-galactosidase were assayed 24 h following transfection, as previously described by McGirt et al. (2014).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Anumanthan G, Halder SK, Osada H, Takahashi T, Massion PP, Carbone DP, Datta PK. 2005. Restoration of TGF-β signalling reduces tumorigenicity in human lung cancer cells. *Br J Cancer* **93:** 1157–1167.

Arrate MP, Vincent T, Odvody J, Kar R, Jones SN, Eischen CM. 2010. MicroRNA biogenesis is required for Myc-induced B-cell lymphoma development and survival. *Cancer Res* **70:** 6083–6092.

Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ. 2010. Development of the human cancer microRNA network. *Silence* **1:** 6.

Barabasi AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5:** 101–113.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

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.

Biswas S, Trobridge P, Romero-Gallo J, Billheimer D, Myeroff LL, Willson JK, Markowitz SD, Grady WM. 2008. Mutational inactivation of TGFBR2 in microsatellite unstable colon cancer arises from the cooperation of genomic instability and the clonal outgrowth of transforming growth factor β resistant cells. *Genes Chromosomes Cancer* **47:** 95–106.

Crawford M, Batte K, Yu L, Wu X, Nuovo GJ, Marsh CB, Otterson GA, Nana-Sinkam SP. 2009. MicroRNA 133B targets pro-survival molecules MCL-1 and BCL2L2 in lung cancer. *Biochem Biophys Res Commun* **388:** 483–489.

Dempke WC, Suto T, Reck M. 2010. Targeted therapies for non-small cell lung cancer. *Lung Cancer* **67:** 257–274.

Dews M, Fox JL, Hultine S, Sundaram P, Wang W, Liu YY, Furth E, Enders GH, El-Deiry W, Schelter JM, et al. 2010. The myc-miR-17~92 axis blunts TGFβ signaling and production of multiple TGFβ-dependent antiangiogenic factors. *Cancer Res* **70:** 8233–8246.

Dutta B, Pusztai L, Qi Y, Andre F, Lazar V, Bianchini G, Ueno N, Agarwal R, Wang B, Shiang CY, et al. 2012. A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. *Br J Cancer* **106:** 1107–1116.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210.

Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin JB. 2010. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Res* **38:** e120.

Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, et al. 2007. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci* **104:** 15805–15810.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92–105.

Gonzalez CR, Matzkin ME, Frungieri MB, Terradas C, Ponzio R, Puigdomenech E, Levalle O, Calandra RS, Gonzalez-Calvar SI. 2010. Expression of the TGF-β1 system in human testicular pathologies. *Reprod Biol Endocrinol* **8:** 148.

Guan P, Yin Z, Li X, Wu W, Zhou B. 2012. Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. *J Exp Clin Cancer Res* **31:** 54.

Guo AY, Sun J, Jia P, Zhao Z. 2010. A novel microRNA and transcription factor mediated regulatory network in schizophrenia. *BMC Syst Biol* **4:** 10.

Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, et al. 2012. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489:** 519–525.

Henley JS, Richards TB, Underwood MJ, Sunderam CR, Plescia M, McAfee TA. 2014. Lung cancer incidence trends among men and women—United States, 2005–2009. *MMWR Morb Mortal Wkly Rep* **63:** 1–5.

Jackson A, Linsley PS. 2010. The therapeutic potential of microRNA modulation. *Discov Med* **9:** 311–318.

Jeon HS, Jen J. 2010. TGF-β signaling and the role of inhibitory Smads in non-small cell lung cancer. *J Thorac Oncol* **5:** 417–419.

Jusufovic E, Rijavec M, Keser D, Korosec P, Sodja E, Iljazovic E, Radojevic Z, Kosnik M. 2012. let-7b and miR-126 are down-regulated in tumor tissue and correlate with microvessel density and survival outcomes in non-small-cell lung cancer. *PLoS One* **7:** e45577.

Kalluri R, Weinberg RA. 2009. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119:** 1420–1428.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28:** 27–30.

Kang HW, Crawford M, Fabbri M, Nuovo G, Garofalo M, Nana-Sinkam SP, Friedman A. 2013. A mathematical model for microRNA in lung cancer. *PLoS One* **8:** e53663.

Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. 2003. MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31:** 3576–3579.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006.

Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, Chang TC, Vivekanandan P, Torbenson M, Clark KR, et al. 2009. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **137:** 1005–1017.

Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and reproducible? *PLoS Comput Biol* **7:** e1001057.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Libermann TA, Zerbini LF. 2006. Targeting transcription factors for cancer gene therapy. *Curr Gene Ther* **6:** 17–33.

Liu Y. 2010. New insights into epithelial-mesenchymal transition in kidney fibrosis. *J Am Soc Nephrol* **21:** 212–222.

Liu Q, Tan Y, Huang T, Ding G, Tu Z, Liu L, Li Y, Dai H, Xie L. 2010. TF-centered downstream gene set enrichment analysis: inference of causal regulators by integrating TF-DNA interactions and protein post-translational modifications information. *BMC Bioinformatics* **11** (Suppl 11): S5.

Ma L, Huang Y, Zhu W, Zhou S, Zhou J, Zeng F, Liu X, Zhang Y, Yu J. 2011. An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. *PLoS One* **6:** e26502.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34:** D108–D110.

McGirt LY, Adams CM, Baerenwald DA, Zwerner JP, Zic JA, Eischen CM. 2014. miR-223 regulates cell growth and targets proto-oncogenes in mycosis fungoides/cutaneous T-Cell lymphoma. *J Invest Dermatol* **134:** 1101–1107.

Munoz NM, Upton M, Rojas A, Washington MK, Lin L, Chytil A, Sozmen EG, Madison BB, Pozzi A, Moon RT, et al. 2006. Transforming growth factor β receptor type II inactivation induces the malignant transformation of intestinal neoplasms initiated by Apc mutation. *Cancer Res* **66:** 9837–9844.

Naeem H, Kuffner R, Zimmer R. 2011. MIRTFnet: analysis of miRNA regulated transcription factors. *PLoS One* **6:** e22519.

Namlos HM, Meza-Zepeda LA, Baroy T, Ostensen IH, Kresse SH, Kuijjer ML, Serra M, Burger H, Cleton-Jansen AM, Myklebost O. 2012. Modulation of the osteosarcoma expression phenotype by microRNAs. *PLoS One* **7:** e48086.

Navarro A, Marrades RM, Vinolas N, Quera A, Agusti C, Huerta A, Ramirez J, Torres A, Monzo M. 2009. MicroRNAs expressed during lung cancer development are expressed in human pseudoglandular lung embryogenesis. *Oncology* **76:** 162–169.

Nyati S, Schinske K, Ray D, Nyati MK, Ross BD, Rehemtulla A. 2011. Molecular imaging of TGFβ-induced Smad2/3 phosphorylation reveals a role for receptor tyrosine kinases in modulating TGFβ signaling. *Clin Cancer Res* **17:** 7424–7439.

Nymark P, Guled M, Borze I, Faisal A, Lahti L, Salmenkivi K, Kettunen E, Anttila S, Knuutila S. 2011. Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. *Genes Chromosomes Cancer* **50:** 585–597.

Palla G, Derenyi I, Farkas I, Vicsek T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435:** 814–818.

Pino MS, Kikuchi H, Zeng M, Herraiz MT, Sperduti I, Berger D, Park DY, Iafrate AJ, Zukerberg LR, Chung DC. 2010. Epithelial to mesenchymal transition is impaired in colon cancer cells with microsatellite instability. *Gastroenterology* **138:** 1406–1417.

Poos K, Smida J, Nathrath M, Maugg D, Baumhoer D, Korsching E. 2013. How microRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation. *PLoS Comput Biol* **9:** e1003210.

Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JM, Macdonald J, Thomas D, Moskaluk C, Wang Y, et al. 2006. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* **66:** 7466–7472.

Raponi M, Dossey L, Jatkoe T, Wu X, Chen G, Fan H, Beer DG. 2009. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res* **69:** 5776–5783.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140.

Schroder MS, Culhane AC, Quackenbush J, Haibe-Kains B. 2011. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **27:** 3206–3208.

Setty M, Helmy K, Khan AA, Silber J, Arvey A, Neezen F, Agius P, Huse JT, Holland EC, Leslie CS. 2012. Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. *Mol Syst Biol* **8:** 605.

Siegel R, Naishadham D, Jemal A. 2013. Cancer statistics, 2013. *CA Cancer J Clin* **63:** 11–30.

Sohler F, Zimmer R. 2005. Identifying active transcription factors and kinases from expression data using pathway queries. *Bioinformatics* **21** (Suppl 2): i115–ii122.

Subramanyam D, Lamouille S, Judson RL, Liu JY, Bucay N, Derynck R, Blelloch R. 2011. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat Biotechnol* **29:** 443–448.

Sun J, Gong X, Purow B, Zhao Z. 2012. Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma. *PLoS Comput Biol* **8:** e1002488.

Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, Homer R, Brown D, Bader AG, Weidhaas JB, et al. 2010. Regression of murine lung tumors by the let-7 microRNA. *Oncogene* **29:** 1580–1587.

Vosa U, Vooder T, Kolde R, Vilo J, Metspalu A, Annilo T. 2013. Meta-analysis of microRNA expression in lung cancer. *Int J Cancer* **132:** 2884–2893.

Wang P, Lushnikova T, Odvody J, Greiner TC, Jones SN, Eischen CM. 2008. Elevated Mdm2 expression induces chromosomal instability and confers a survival and growth advantage to B cells. *Oncogene* **27:** 1590–1598.

Wang L, Jia P, Wolfinger RD, Chen X, Grayson BL, Aune TM, Zhao Z. 2011. An efficient hierarchical generalized linear mixed model for pathway analysis of genome-wide association studies. *Bioinformatics* **27:** 686–692.

Wernicke S. 2006. Efficient detection of network motifs. *IEEE/ACM Trans Comput Biol Bioinform* **3:** 347–359.

Wernicke S, Rasche F. 2006. FANMOD: a tool for fast network motif detection. *Bioinformatics* **22:** 1152–1153.

Xu T, Liu X, Han L, Shen H, Liu L, Shu Y. 2014. Up-regulation of miR-9 expression as a poor prognostic biomarker in patients with non-small cell lung cancer. *Clin Transl Oncol* **16:** 469–475.

Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, et al. 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9:** 189–198.

Yang G, Yang X. 2010. Smad4-mediated TGF-β signaling in tumorigenesis. *Int J Biol Sci* **6:** 1–8.

Yu H, Greenbaum D, Xin Lu H, Zhu X, Gerstein M. 2004. Genomic analysis of essentiality within protein networks. *Trends Genet* **20:** 227–231.

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.

Zhao M, Sun J, Zhao Z. 2013. Synergetic regulatory networks mediated by oncogene-driven microRNAs and transcription factors in serous ovarian cancer. *Mol Biosyst* **9:** 3187–3198.

Zheng S, Tansey WP, Hiebert SW, Zhao Z. 2011. Integrative network analysis identifies key genes and pathways in the progression of hepatitis C virus induced hepatocellular carcinoma. *BMC Med Genomics* **4:** 62.