Discover prognostic RNAs through identifying lncRNA-miRNA-mRNA triplets in lung squamous cell carcinoma

Xue Yang[1], Zhuang Yu[1,*], Ning An[1,*]

[1]Department of Oncology, the Affiliated Hospital of Qingdao University, Qingdao, Shandong, 266003, China

[*]These authors contributed equally to this work

Correspondence to: Ning An, **email:** nick_anning@sina.com
Zhuang Yu, **email:** yuzhuang2002@163.com

Keywords: *lung squamous cell carcinoma, triplet regulation, biological network, prognostic genes*

Received: *March 04, 2017* ***Accepted:*** *April 11, 2017* ***Published:*** *April 19, 2017*

ABSTRACT

Carcinogenesis involves multi-level RNA dysregulations, including long non-coding RNA, microRNA and mRNA. Therefore, only looking into one RNA level of cancer is not sufficient to uncover the intricate underlying mechanisms. In this study, an integrative strategy was conducted to systematically analyze the characterized molecular patterns of lung squamous cell carcinoma. Paired samples (cancer and adjacent normal tissues) in each RNA level were retrieved to identify differentially expressed RNAs. The regulatory triplets of lncRNA-microRNA-mRNA were identified after considering dysregulation pattern and evidence of publicly available databases. Thus, 32 significant mRNAs were retrieved by means of random walk algorithm. In four independent datasets, Kaplan–Meier survival and Cox regression analyses both confirmed that the expression of these genes was significantly associated with overall survival of LUSC patients.

INTRODUCTION

Lung cancer holds the largest mobility and mortality rate among all types of cancer worldwide, and more than 1.8 million new cases emerged every year (13% of the new cancer cases) [1]. The same ranking of lung cancer was also confirmed in China [2]. Lung squamous cell carcinoma (LUSC) accounts up to 30% of lung cancer [3]. Up to date, Cisplatin plus gemcitabine is still the first-line treatment of advanced LUSC [4], and therapeutic options for LUSC patients remain scarce because no molecular targets have ever been identified [5].

Carcinogenesis is an extremely complex process, of which the initiation and progression is closely associated with not only the protein coding mRNAs, but also non-coding RNAs composing up to ~ 98% of the human genome [6]. Long non-coding RNAs (lncRNAs), of which the length is greater than 200 nucleotides (nt) but appearing to lack protein-coding potential, are produced through complex transcription interwoven between and within protein-coding genes [7]. Emerging evidence repetitively addressed that lncRNA might wield substantial weight upon cancer transformation [8–10], for instance, in post-transcriptional regulation [11], tumorigenesis [12], and embryonic development [13]. microRNAs (miRNAs) are a class of small non-coding RNA (~ 22 nt) that regulate gene expression by binding to the 3'-untranslated region (3'-UTR) of target genes, leading to the degradation or protein translation inhibition of target genes [14], and predicted to be the regulator of more than 60% of all protein-coding genes in mammal [15] and wield its regulatory power almost across every aspect of cellular process [16, 17]. The important role played by non-coding RNAs indicated that while looking into the dysregulation of mRNA expression in cancer, the aberrant patterns of multi-level events, including lncRNA and miRNA, should also be paid considerable attention. Therefore, the integrative analysis of LUSC multi-level RNA is urgently needed.

Among the hypotheses for generalized mechanisms of lncRNA function, the one receiving the most notable attention is the competitive endogenous RNA (ceRNA) hypothesis. The ceRNA hypothesis posits that specific lncRNAs can inhibit miRNA activity through sequestration, thereby up-regulating mRNA expression level [17]. For example, lncRNA *HULC* could reduce both *miR-372* expression and activity leads to reducing translational repression of its target gene, *PRKACB*, which in turn induces phosphorylation of *CREB* in liver cancer [16]. In clear-cell renal cell carcinoma, lncRNA *PTENP1* functions as a ceRNA of *PTEN* for *miR-21* to suppress

cancer progression, and the expression of *PTENP1* and *PTEN* is positively correlated, and the expression of both is inversely correlated with *miR-21* expression [18]. Therefore, the negative correlation between miRNA and lncRNA or mRNA could be used as a filtering rule to identify lncRNA-miRNA-mRNA regulatory triplet.

RESULTS

The schematic of this study is illustrated in Figure 1.

Identification of differentiated lncRNAs, miRNAs and mRNAs according to paired TCGA data

Differentially expressed RNAs (DERs) in LUSC were composed of 1,223 up-regulated and 1,177 down-regulated lncRNAs, 228 up-regulated and 81 down-regulated miRNAs, and 4,157 up-regulated and 5,073 down-regulated mRNAs (Figure 2A–2C). In addition, principle component analysis (PCA) of the TCGA paired data indicated that human lung carcinogenesis was characterized by substantial changes in transcriptomic features of three RNA levels (Figure 2D–2F). The cancer samples and normal samples could be clustered into two distinct groups, respectively. Therefore, we collected DERs for further analysis to reduce signal noise. GO analysis for differentially expressed mRNAs was conducted via the DAVID bioinformatics tool (http://david.abcc.ncifcrf.gov/). The result of GO enrichment analysis indicated that the down-regulated mRNAs were functionally concentrated on "immune response", since the majority of the enriched GO terms were the offspring of this GO term (*FDR*<0.001, Figure 2G, Supplementary Table 2), while up-regulated mRNAs were functionally concentrated on "cell cycle" (*FDR*<0.001, Figure 2H, Supplementary Table 3), which was also confirmed by our previous research [19].

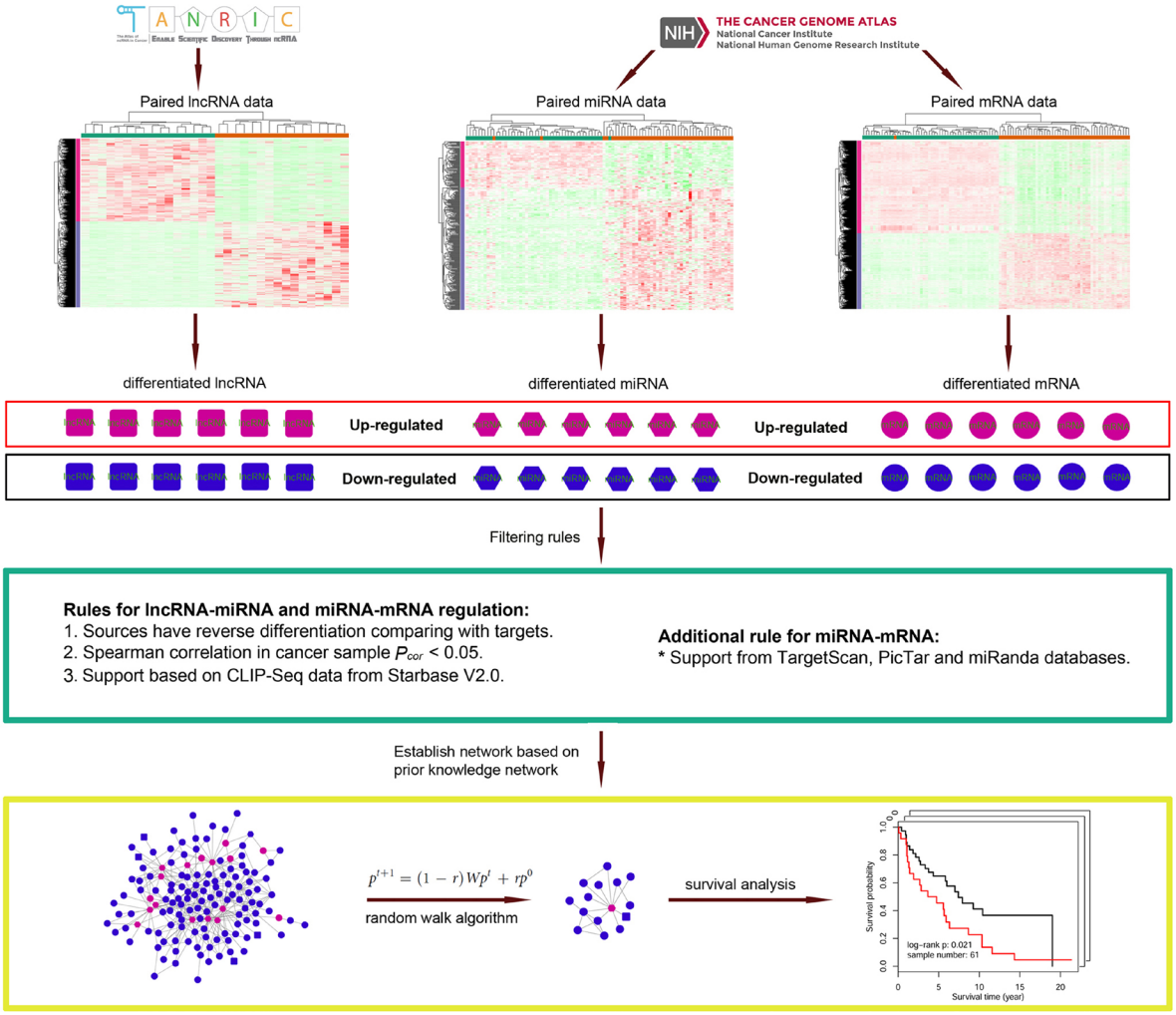

Figure 1: Schematic of methodology applied in this study. Step I: Using TCGA paired LUSC samples to identify differentially expressed RNA in three levels; Step II: Identification of lncRNA-miRNA-mRNA regulatory triplets according to several filtering rules; Step III: Identification of mRNAs significantly affected by the dysregulation of upstream lncRNAs and miRNAs, by applying random walk with restart algorithm in the merged network. Step IV: Survival analysis of identified significant RNAs to evaluate their prognostic value.

Establish lncRNA-miRNA-mRNA triplet regulatory relation

For lncRNA-miRNA regulation, filtering rules consisted of reverse differentiation direction, intra-LUSC Spearman correlation and support from Starbase V2.0 database, and finally 41 lncRNAs were found regulating 42 miRNAs; as for miRNA-mRNA regulation, with the additional requirement of three miRNA-mRNA sequence algorithms, finally we found 134 miRNAs regulating 2411 mRNAs.

Random walk in HPRD-KEGG merged biological network

We then projected these regulatory triplets onto HPRD-KEGG merged biological network, and used the biggest connected component (BCC, containing 540 mRNAs regulated by 41 lncRNAs and 41 miRNAs, Figure 3) as the walking graph for random walk. mRNAs with triplet regulation in the BCC were all used as source nodes until the possibility vector reached steady-state. Source nodes were weighted according to the aforementioned formula. The initial probability vector p_0 was obtained by normalizing the score vector (n = 540) so that the sum of the vector is equal to 1 (the input of random walk algorithm). When the steady-state was finally reached, all the genes in the BCC were scored with p^{∞} (n = 540, output of random walk algorithm), and thus the mRNAs with significantly high score were mostly affected by upstream lncRNA and miRNA dysregulations. Therefore, 32 significant mRNAs ($p < 0.02$) in respect to steady-stage probability were collected through 10,000 permutations (Figure 4), and algorithmically these genes received the most influence imposed by upstream dysregulation of 34 lncRNAs and 27 miRNAs

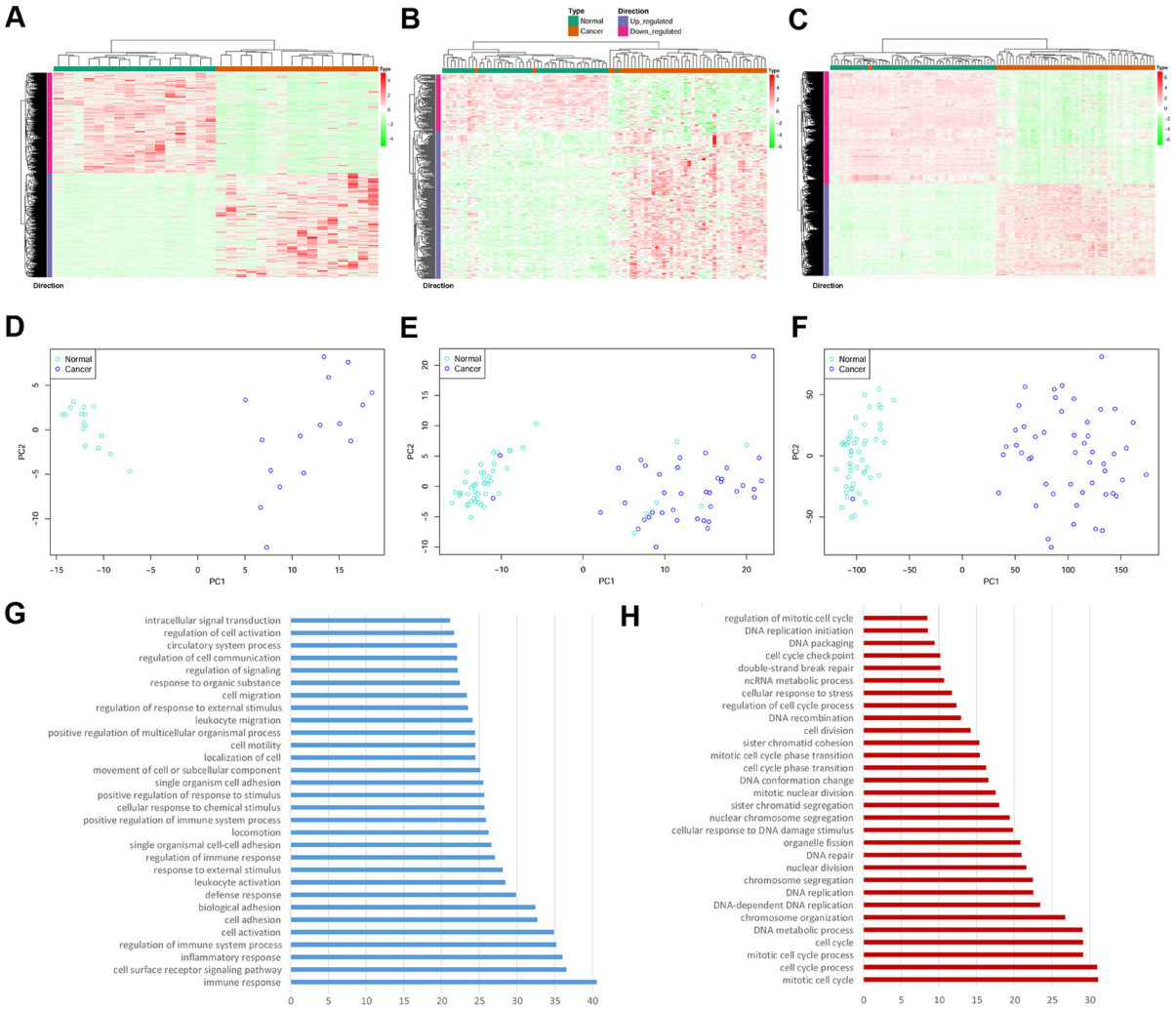

Figure 2: Identification of differentially expressed lncRNAs, miRNAs and mRNAs using TCGA paired samples. (**A–C**) Heat maps of differentially expressed lncRNAs, miRNAs and mRNAs in corresponding paired TCGA data, respectively. (**D–F**) PCA analyses with differentially expressed lncRNAs, miRNA, and mRNAs discriminating distinct molecular characteristics between cancer and adjacent normal tissues, respectively. (**G**) GO enrichment analysis of down-regulated mRNAs in LUSC. The length of the blue bars is equal to log2 transformed FDR value of GO analysis. (**H**) GO enrichment analysis of up-regulated mRNA in LUSC.

(Supplementary Table 4). The Circos plot of these triplet regulations were illustrated in Figure 5. The heatmaps of upstream 34 lncRNAs, 27 miRNAs and 32 significant mRNAs were shown in Figure 6A with 158 TCGA LUSC samples, and corresponding clinicopathological factors were also illustrated.

Validation of significant genes' prognostic value via survival analysis

We search GEO database and used TCGA LUSC, GSE73403 (our samples), GSE30219 and GSE11969 to evaluate the prognostic value of these 32 significant mRNAs. Kaplan–Meier survival analysis indicated that these 32 significant mRNAs performed well in all four independent cohorts (TCGA, n = 483, p = 0.043; GSE73403, n = 69, p = 0.024; GSE11969, n = 35, p = 0.008, GSE30219, n = 61, p = 0.021, Figure 6B). The prognostic values of upstream lncRNAs and miRNAs were also evaluated using TCGA datasets, indicating the expression of upstream lncRNAs was significantly associated with OS (n = 212, p = 0.009, Figure 7A–7B), rather than that of upstream miRNAs (n = 458, p = 0.350, Figure 7C–7D).

Confirmation of the prognostic value of the three-level RNAs

The Cox proportional hazards regression model was used to evaluate the independence of the prognostic

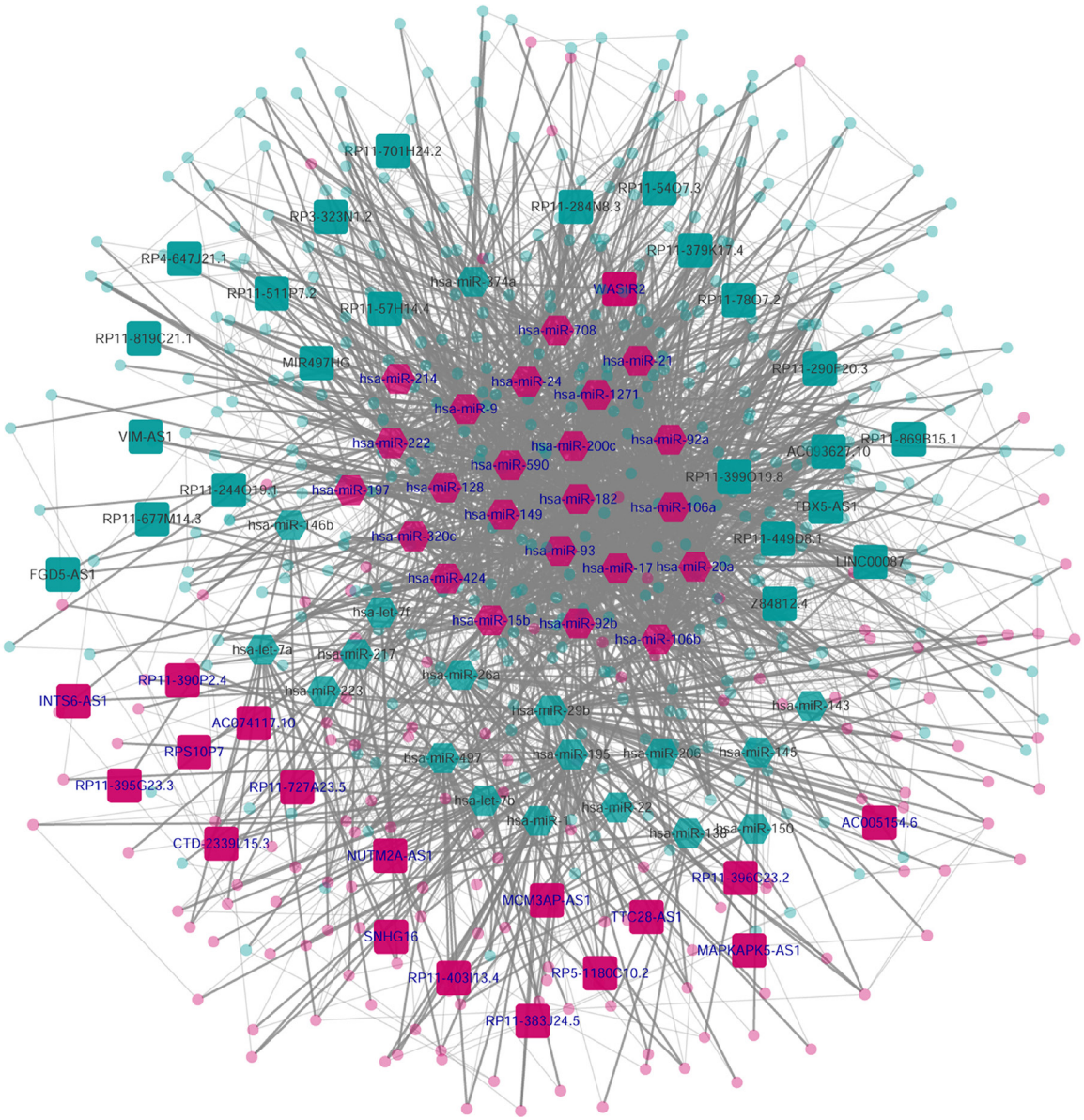

Figure 3: Biggest connected component (BCC) for random walk. BCC includes 540 mRNAs, 41 lncRNAs and 41 miRNAs. Round dots represented mRNAs, hexon represented miRNAs, and square represented lncRNAs. Up-regulated RNAs were colored red, and down-regulated ones were dark green.

factors in a stepwise manner (Table 1). In each independent cohort, samples with OS, age, sex, American Joint Committee on Cancer (AJCC) stage information and smoking index were used to perform the analysis. The expressions of these 32 mRNAs, 34 upstream lncRNAs [hazard ratio (HR): 1.596; 95% confidence interval (CI): 1.051~2.424; $p = 0.028$], rather than 27 miRNAs, were also confirmed as an independent prognostic factor in predicting patients' OS, respectively. (Table 1)

DISCUSSION

The booming amount of high-throughput and multi-dimensional genomic data usher us into a new era, when the tremendously complicated molecular mechanism of carcinogenesis could be possibly perceived and dissected in a more integrative perspective. Alterations in the primary or secondary structure, and expression levels of lncRNAs as well as their cognate RNA-binding proteins are often associated with human diseases [20]. In the last decade, the hypothesis of ceRNA was putatively accepted as a novel layer of gene regulation. Emerging evidences have indicated that lncRNA functioned as miRNA sponges by decoying miRNAs from other target transcripts, and therefore wielded great weight upon the process of carcinogenesis. For instance, lncRNA *loc285194* was proven as a *p53*-regulated tumor suppressor through reciprocal through repression the expression level of *miR-211* [21]. Increased lncRNA *CYP4Z2P* and *CYP4Z1*-3′UTR expression promotes tumor angiogenesis in breast cancer partly via miRNA-dependent activation of *PI3K*/*Akt* and *ERK*1/2 by inhibiting several miRNAs

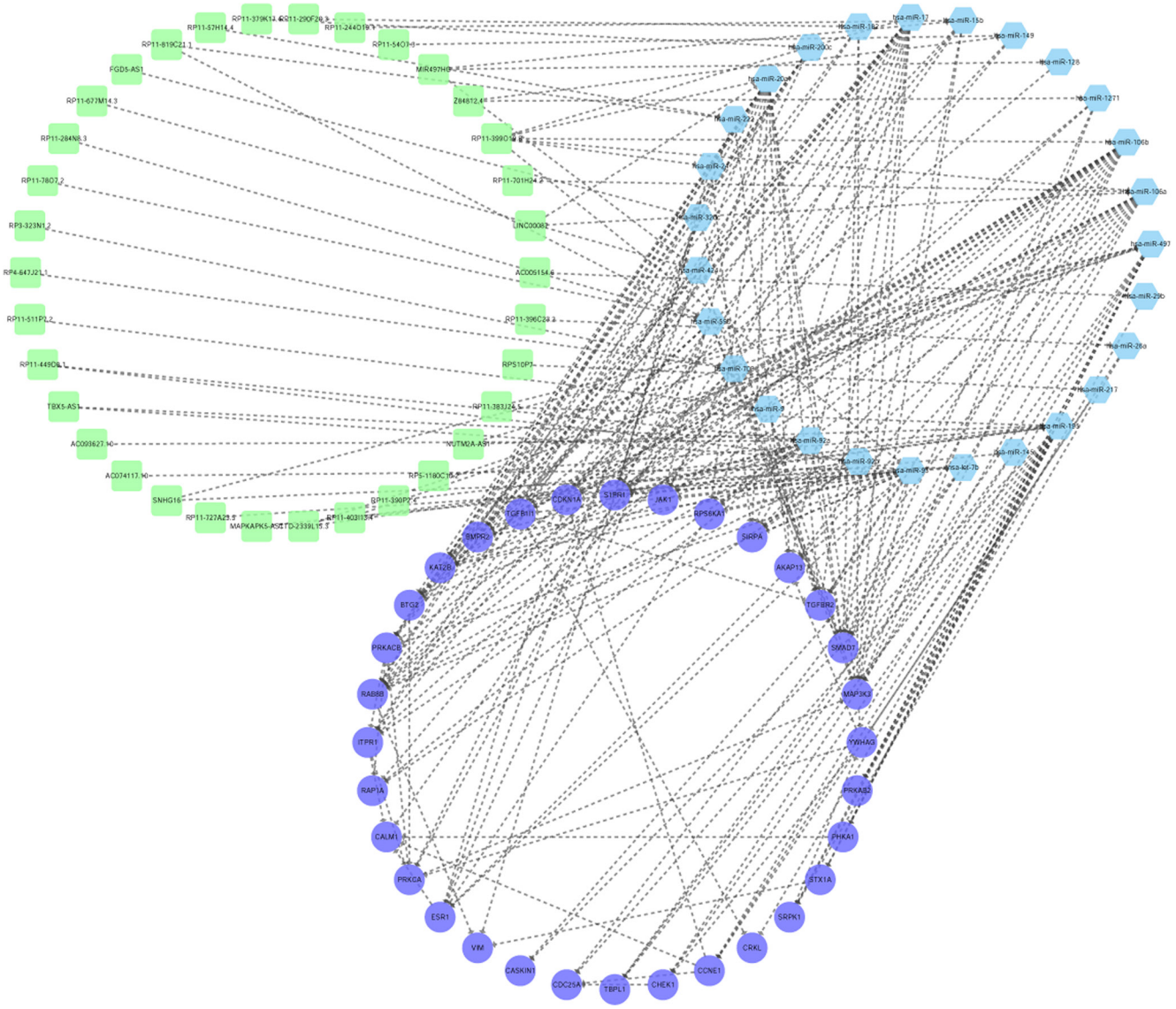

Figure 4: Subnetwork composed of 32 significant mRNAs, 34 upstream lncRNAs and 27 miRNAs. The layout was set as circular format. Purple round dots represented mRNAs, light blue hexons represented miRNAs, and light green squares represented lncRNAs.

[22]. However, our knowledge about the underlying mechanism of lncRNA acting as ceRNAs in LUSC is still limited. In this study, we systematically analyzed LUSC's RNA sequencing data, including lncRNA, miRNA and mRNA expression level, to discover novel and important molecular dysregulations in a more comprehensive manner, shedding significant light upon the novel layer of post-transcriptional regulation in human cancers.

In order to identify lncRNA-miRNA-mRNA regulatory triplet of great importance with as much accuracy as possible, we first identify differentially expressed RNAs in three levels based on TCGA paired data. According to the ceRNA hypothesis, under the most circumstances, lncRNA and mRNA should be negatively correlated with miRNA expression level, since lncRNAs may function as miRNA sponges, to downregulate the

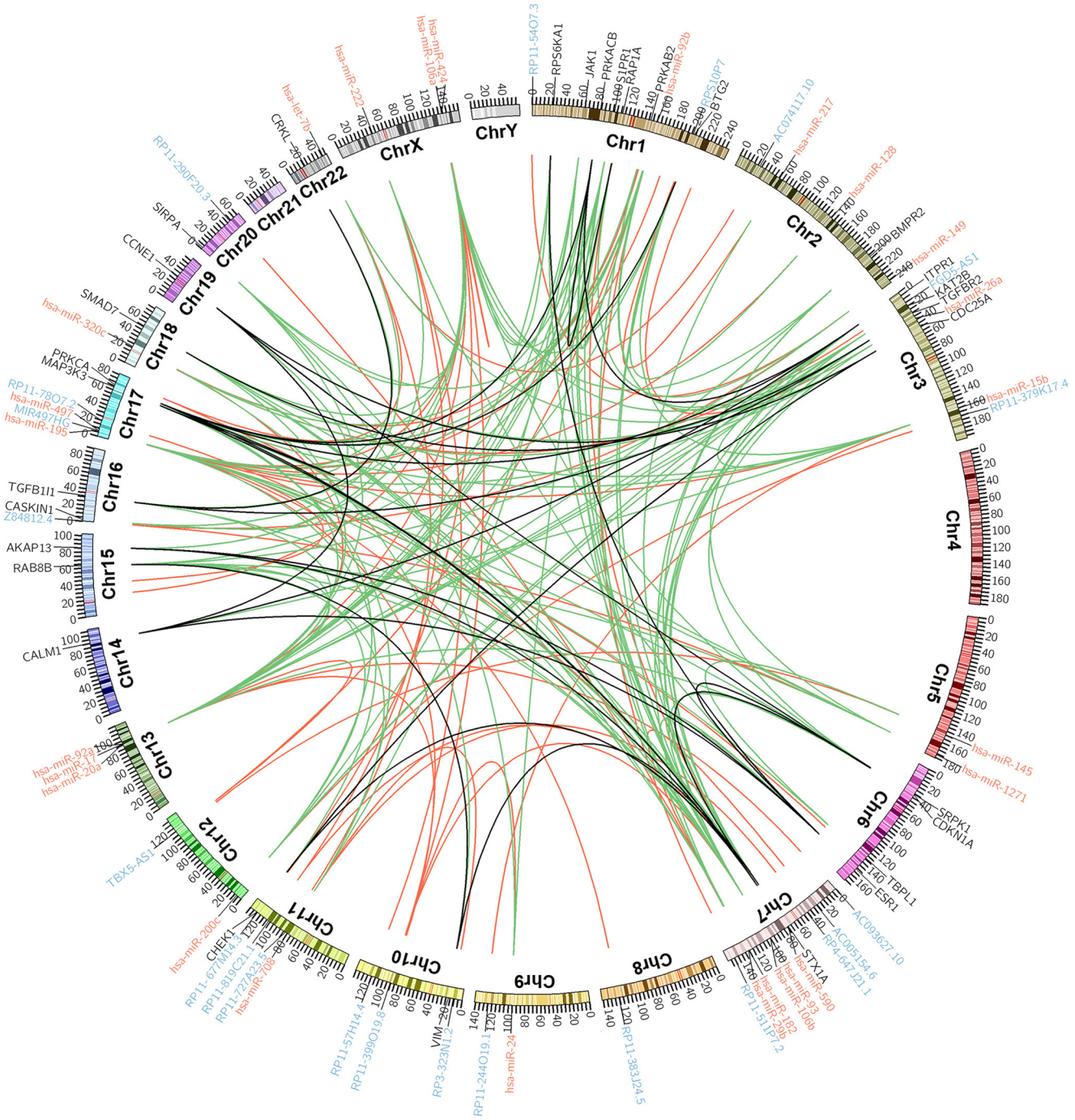

Figure 5: Circos plot illustration of lncRNA-miRNA-mRNA triplet regulation. An ideogram of a normal karyotype is shown in the outermost ring. Black labels represented 32 significant mRNAs. Blue labels represented 34 lncRNAs. Red labels represented 27 miRNAs. Black links represented mRNA-mRNA interactions. Red links represented lncRNA-miRNA regulation. Green links represented miRNA-mRNA regulation.

expression and activities of miRNAs, thereby modulating the de-repression of miRNA targets and imposing an additional level of post-transcriptional regulation [23]. Therefore, we searched for the regulatory pairs, of which source nodes had the reverse differentiation direction in comparison to target nodes, and were negatively correlated with target nodes. Starbase database contained invaluable information for deciphering lncRNA-miRNA interactions, miRNA-mRNA interactions and ceRNA networks based on 108 CLIP-Seq datasets, whereby the involvement of Starbase evidence, could greatly consolidate the credibility of the regulatory relation.

GO analysis indicated that differentially expressed mRNAs were involved in immune response and cell cycle processes. Tumor-associated cell cycle defects may induce aberrant proliferation as well as genomic and chromosomal instability, which are often mediated by alterations in cyclin-dependent kinase (*CDK*) activity [24]. Infection and chronic inflammation contribute to an estimated 25% of all cancers worldwide [25]. In developmental biology, the fetus, which in many ways behaves like an allogenic transplant, also evades maternal immune-surveillance through mechanisms similar to those observed in tumors [26]. Indeed, excessive proliferation (activation of cell cycle genes) and immune-surveillance evasion (suppression of immune genes) allow tumors to obtain territorial expansion advantages compared to normal cells.

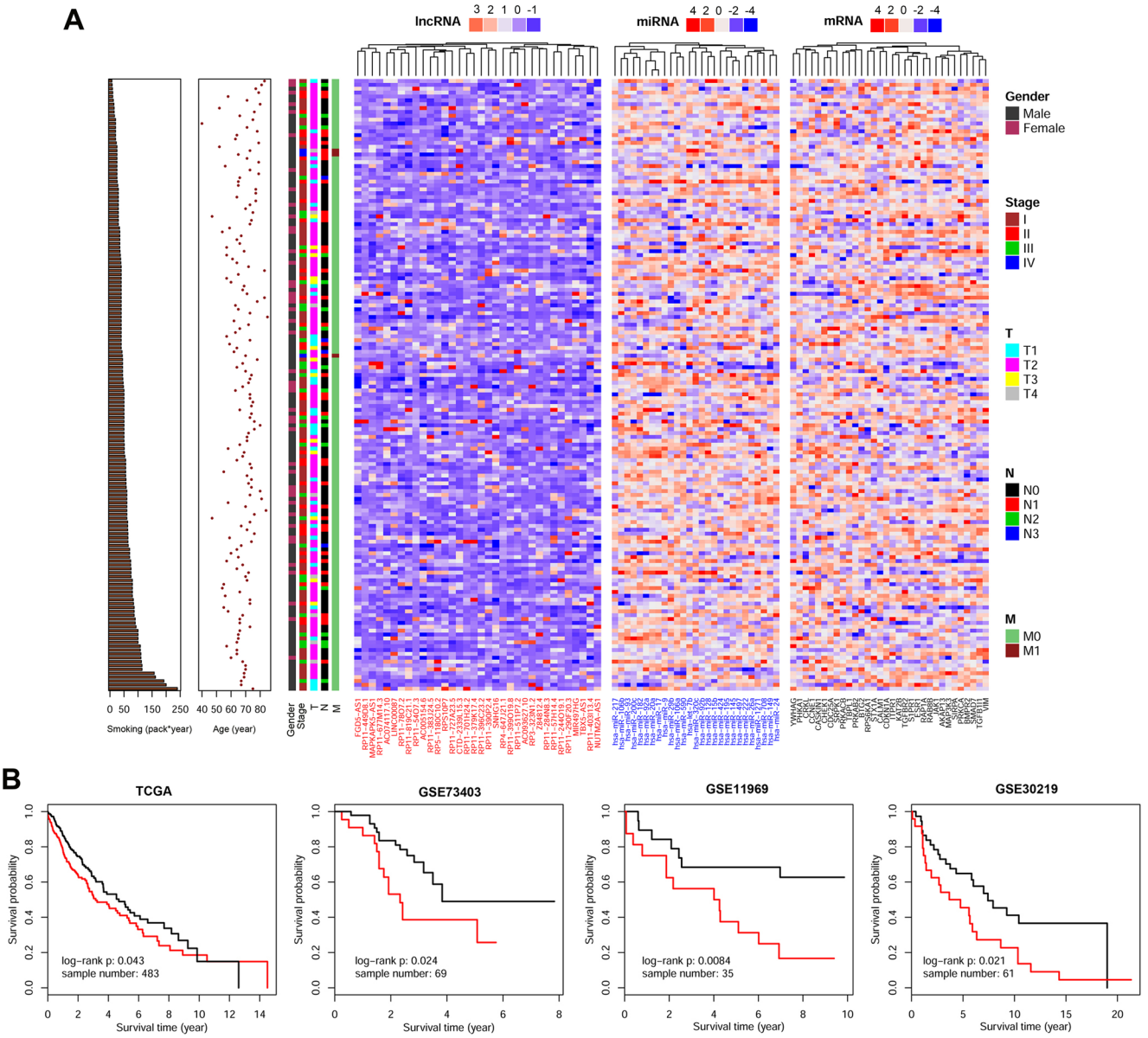

Figure 6: Heatmaps of three RNA levels and survival analysis of significant mRNAs in four independent datasets. (**A**) Heatmaps of 34 lncRNAs, 27 miRNAs and 32 significant mRNAs using 158 TCGA LUSC samples. Row represented 158 TCGA samples, ordered by smoking index. (**B**) Kaplan–Meier survival analysis of 32 significant mRNAs in four independent data sets. Survival analysis was performed to discriminate OS between k-means assigned groups.

Table 1: Univariate and multivariate analyses of overall survival (Cox proportional hazards regression model)

Factors	Univariate Cox regression		Multivariate Cox regression	
	HR (95% CI)	*P*	HR (95% CI)	*P*
GSE73403 (mRNA, n = 69)				
Age	0.999 (0.959~1.040)	0.950	-	-
Sex (Male/Female)	1.125 (0.264~4.789)	0.873	-	-
Stage (III/I+II)	2.283 (1.049~4.969)	**0.037**	2.746 (1.231~6.126)	**0.014**
Smoking (pack-year)	1.008 (0.995~1.021)	0.210	-	-
K-means[a]	2.215 (1.034~4.746)	**0.041**	2.658 (1.209~5.844)	**0.015**
GSE30219 (mRNA, n = 61)				
Age	1.038 (1.004~1.073)	**0.028**	1.058 (1.020~1.097)	**0.003**
Sex (Male/Female)	0.438 (0.153~1.256)	0.125	-	-
Stage (II+III/I)	2.202 (1.074~4.512)	**0.031**	3.698 (1.657~8.251)	**0.001**
K-means[a]	2.011 (1.103~3.665)	**0.023**	2.210 (1.201~4.067)	**0.011**
GSE11969 (mRNA, n = 35)				
Age	1.030 (0.975~1.088)	0.295	-	-
Sex (Male/Female)	1.282 (0.171~9.592)	0.809	-	-
Stage (III+IV/I+II)	3.407 (1.395~8.323)	**0.007**	2.864 (1.157~7.090)	**0.023**
K-means[a]	3.356 (1.300~8.663)	**0.012**	2.835 (1.077~7.465)	**0.035**
TCGA (mRNA, n = 480)				
Age	1.017 (0.999~1.035)	0.057	-	-
Sex (Male/Female)	1.109 (0.797~1.542)	0.539	-	-
Stage (II + III + IV/I)	1.391 (1.046~1.849)	**0.023**	1.391 (1.046~1.849)	**0.023**
Smoking (pack-year)	1.000 (0.996~1.005)	0.905	-	-
K-means[a]	1.357 (1.020~1.806)	**0.036**	1.358 (1.021~1.807)	**0.036**
TCGA (lncRNA, n = 212)				
Age	1.036 (1.010~1.063)	**0.006**	1.039 (1.012~1.067)	**0.004**
Sex (Male/Female)	1.416 (0.886~2.264)	0.146	-	-
Stage (II + III + IV/I)	1.602 (1.070~2.397)	**0.013**	1.621 (1.079~2.436)	**0.020**
Smoking (pack-year)	1.000 (0.994~1.007)	0.924	-	-
K-means[a]	1.357 (1.020~1.806)	**0.022**	1.596 (1.051~2.424)	**0.028**
TCGA (miRNA, n = 456)				
Age	1.015 (0.997~1.034)	0.105	-	-
Sex (Male/Female)	1.033 (0.735~1.450)	0.853	-	-
Stage (II + III + IV/I)	1.456 (1.081~1.962)	**0.013**	-	-
Smoking (pack-year)	1.000 (0.995~1.005)	0.995	-	-
K-means[a]	1.077 (0.799~1.452)	0.628	-	-

[a]Based on the k-means algorithm to divide samples into two groups. Significant *p* values were in bold ($p < 0.05$). Abbreviations: *HR*, hazard ratio; *CI*, confidence interval.

In this study, we adopted a simple and effective computational strategy to randomly walk through mRNAs with significantly dysregulated lncRNAs and miRNAs in HPRD and KEGG merged biological network. We hypothesized downstream target mRNAs may be conferred with copious prognostic information, and the same strategy was also adopted by previous researches [27, 28]. Therefore, mRNAs with upstream lncRNA-miRNA regulation were collected, and the biggest connected component composed of these mRNAs were used as walking graph for random walk. Random walk with restart was adopted to decipher gene to disease association in prior-knowledge based network, whose performance was proven to be much more superior to other methods, such as neighborhood approaches [29–31]. The advantage of this strategy is that it subtly combines observed triplet

regulation with knowledge-based regulatory network, tracing the information flow which would be greatly accumulated in significant genes. Nodes in the network were weighted with the mean of regulatory consistency, quantifying the authenticity the regulatory pair held. In this way, significant mRNAs identified by random walk through biological network contained specific genes that underwent the considerable dysregulation of upstream lncRNA-miRNA during carcinogenesis.

Many of these significant mRNAs obtained through random walk algorithm were closely related to the initiation and progression of LUSC. For example, *JAK1* is responsible for *STAT3* activation in non-small-cell lung cancer and inactivation of *JAK1-STAT3* can inhibit lung cancer growth [32]. *SRPK1* could promote a stem cell-like phenotype in non-small-cell lung cancer via *Wnt/β*-catenin signaling [33]. Overexpression of *CRKL* promoted cell invasion through upregulation of *MMP9* expression and activation of *ERK* pathway in lung cancer [34]. The abundance of supporting researches further endorsed our study that the differentially expressed genes significantly influenced by the dysregulation of upstream lncRNAs and miRNAs might be extremely crucial during the process of lung carcinogenesis.

Since candidate genes were collected based on aberrant patterns in multi-level of TCGA RNA sequencing data, we used microarray data sets with OS information from GEO database in conjunction with TCGA to test the prognostic value of these significant genes and corresponding upstream regulators. The reason why only mRNAs and lncRNAs, rather than miRNAs in regulatory triplet, were evaluated as valuable in OS prediction is still unclear. Probably miRNAs might be functioned as a cushion mediating the proper expression level for both lncRNA and mRNA, and therefore the two-side tradeoff compromises the prognostic value of sandwiched miRNAs. The underlying mechanism needs further exploration. Moreover, smoking is no doubt the leading hazardous factor, especially for LUSC, confirmed

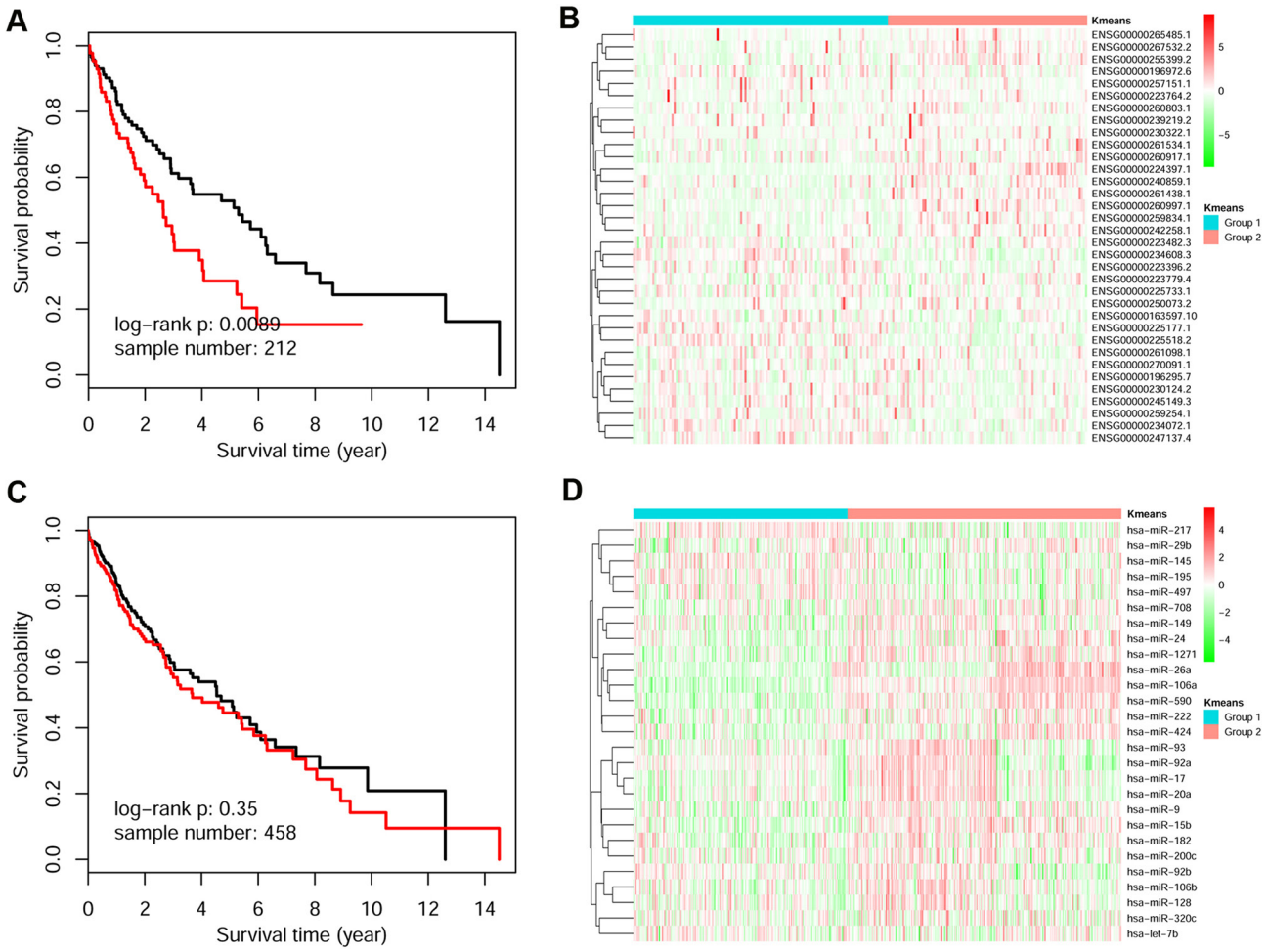

Figure 7: Survival analysis of the upstream lncRNAs and miRNAs in TCGA dataset. (**A**) Kaplan–Meier survival analysis of 34 upstream lncRNAs in the TCGA data, in which patients are divided into two k-means assigned groups. (**B**) The heatmap of 34 lncRNAs in the TCGA data. Rows represent lncRNAs, which were clustered using an unsupervised clustering algorithm, while columns represent samples, which are divided into two groups according to k-means algorithm. (**C–D**) the survival analysis and heatmap of 27 upstream miRNAs in TCGA dataset.

by numerous researches [35]. However, at least in TCGA database and our microarray data, smoking index was not significantly correlated with LUSC patient's OS (Table 1). In Figure 6A, 158 TCGA LUSC samples were ordered by smoking index, and all the listed clinicopathological factors and the expression pattern of three RNA levels seemed randomly scattered, suggesting smoking works as the ignitor for LUSC formation, but probably not as the indicator for LUSC's prognosis and progression.

MATERIALS AND METHODS

Data retrieval

The multi-dimensional data of LUSC associated datasets were retrieved from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/). Two levels of RNA sequencing data [RNASeq by Expectation Maximization (RSEM) normalized read counts] were downloaded, including 51 pairs of mRNA sequencing level 3 data (including 20,531 mRNAs) and 45 pairs of miRNA sequencing level 3 data (including 546 mature miRNAs processed by TCGA). LncRNA sequencing data of LUSC was downloaded from The Atlas of ncRNA in Cancer database (TANRIC, http://ibl.mdanderson.org/tanric/_design/basic/index.html), including 16 pairs of LUSC samples and 12,727 lncRNAs. The RNAs with 0 read count in more than half of corresponding samples were excluded. The expression value was log2 transformed to approximate normal distribution. TCGA LUSC samples with overall survival (OS) information (n = 483) were used for prognostic evaluation. Datasets with OS information (GSE30219 and GSE11969, only LUSC samples were included) were downloaded from GEO database. All clinical information was extracted from the original publications.

Microarray expression profiling of LUSC samples

The global mRNA expression profile of 69 LUSC samples with OS information were conducted in our previous research [19] (Supplementary Table 1). After histopathological evaluation and RNA integrity analysis, all these samples were purified and analyzed using Agilent microarrays. Total RNA samples were labeled and hybridized to Agilent 4*44K Whole Human Genome Oligo Microarrays (G4112F). Normalized expression data were extracted with R package "limma" using cyclic loess method, and the ComBat algorithm was utilized to eliminate potential batch effects. The expression levels of 18,453 genes were obtained as the median value of all probes mapping to a particular gene. The raw and processed data have been deposited in GEO database with the series accession numbers GSE73403.

Establishment of lncRNA-miRNA-mRNA triplet regulatory relation

First, significantly differentiated lncRNAs, miRNAs and mRNAs were identified based on paired t test using LUSC RNA sequencing data, and the p value was also adjusted with false discovery rate (FDR) algorithm (FDR<0.05).

Several rules must be obeyed in order to establish lncRNA-miRNA-mRNA regulatory relation (Figure 1), several rules: (i) source nodes must have the reverse differentiation direction comparing with target nodes (i.e., up-regulated lncRNAs regulate down-regulated miRNAs, which regulate up-regulated mRNAs; or down-regulated lncRNAs regulate up-regulated miRNAs, which regulate down-regulated mRNAs; (ii) source nodes must be negatively correlation with target nodes according to Spearman correlation algorithm within cancer samples (p < 0.05); (iii) lncRNA-miRNA and miRNA-mRNA regulation both required support from CLIP-Seq data in Starbase V2.0 database [36]; (iv) miRNA-mRNA regulatory pair was regarded as solid only if they satisfied at least two of three sequence algorithms (miRanda [37], TargetScan [38], PicTar [39]) as well.

Establish merged biological network

The protein–protein interaction network was downloaded from the Human Protein Reference Database (HPRD), and Kyoto Encyclopedia of Genes and Genomes (KEGG) network. Therefore, gene regulatory network was established by merging the two networks, including 10,340 nodes and 60,642 edges after eliminating self-loops and duplicated edges. Therefore, the regulatory network was composed by mRNAs with lncRNA-miRNA regulation and corresponding regulators.

Random walk in merged biological network

Taking advantage of knowledge-based network topology, random walk algorithm was utilized to identify genes algorithmically most affected by corresponding lncRNAs and miRNAs [40]. In the network, genes of interest were designated as information source (i.e., source nodes) and the remaining genes in the network as the information target (i.e., target nodes). Biggest connected component (BCC) of interest is used as the walking graph for random walk. mRNAs with triplet regulation in the BCC were all used as source nodes until the possibility vector reached steady-state. Source nodes were weighted according to following formula:

$$W = \frac{|Cor_{l-mi}| + |Cor_{mi-m}|}{2} \quad (1)$$

$Cor_{l\text{-}mi}$ represented the Spearman correlation coefficient between lncRNA and its regulated miRNA, whereas $Cor_{mi\text{-}m}$ represented that between miRNA and regulated mRNA. Therefore, W denoted the regulatory consistency, i.e., the bigger W is, the more likely this triplet regulatory relation exists. The information flow originates from source nodes iteratively and randomly transmits to their neighbors with a probability proportional to their topological features. At each step, the information can flow back to the source nodes with the same probability. The final steady-state probability assigned to each gene in the network reflects the integrated influence imposed by source nodes combining network topology. Formally, the random walk with restart is defined as:

$$p^{t+1} = (1-r)Wp^t + rp^0 \quad (2)$$

where W is the column-normalized adjacency matrix of network, and p^t is a vector in which the genes in the network holds probability in the iterative process up to step t. Source nodes were weighted with initial probability vector p^0 (the sum of its elements was equal to 1), and r represents restart probability ($r = 0.7$ in this study). All the genes in the network were ranked according to the values in the steady-state probability vector p^∞. This was obtained at query time by performing the iteration until the difference between p^t and p^{t+1} (measured by the L1 norm) was lower than 10^{-10}. In order to obtain genes with significantly high steady-state probability, 10,000 permutations of node labels (with network topology remained the same) were conducted to calculate the null distribution of final probability for each gene. The p value was termed as the ratio of random values that were greater than the observed final probability. Genes with $p < 0.02$ were regarded as the genes significantly afflicted by upstream lncRNA and miRNA dysregulations.

Survival analysis

We clustered LUSC patients based on the expression profile of cancer patients used k-means algorithm. Kaplan–Meier survival analysis and the log-rank test were used to evaluate prognostic differences between the two k-means assigned groups [41–43]. The Cox proportional hazards regression model was used to evaluate the independence of the prognostic factors in a stepwise manner. Samples in each dataset with complete patient age, sex, stage, smoking index and OS information were used for Cox analysis, and a value of $p < 0.05$ was regarded as significant.

Statistical analysis

All the data analyses were conducted using R programming (Version 3.3.1) and Matlab (Version 2015b) language.

CONFLICTS OF INTEREST

None.

REFFERENCES

1. McGuire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. Adv Nutr. 2016; 7:418–9.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin. 2016; 66:115–32.
3. Heist RS, Mino-Kenudson M, Sequist LV, Tammireddy S, Morrissey L, Christiani DC, Engelman JA, Iafrate AJ. FGFR1 amplification in squamous cell carcinoma of the lung. J Thorac Oncol. 2012; 7:1775–80. doi: 10.1097/JTO.0b013e31826aed28.
4. Thatcher N, Hirsch FR, Luft AV, Szczesna A, Ciuleanu TE, Dediu M, Ramlau R, Galiulin RK, Balint B, Losonczy G, Kazarnowicz A, Park K, Schumann C, et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. Lancet Oncol. 2015; 16:763–74.
5. Tiseo M, Gelsomino F, Alfieri R, Cavazzoni A, Bozzetti C, De Giorgi AM, Petronini PG, Ardizzoni A. FGFR as potential target in the treatment of squamous non small cell lung cancer. Cancer Treat Rev. 2015; 41:527–39.
6. Takenaka K, Chen BJ, Modesitt SC, Byrne FL, Hoehn KL, Janitz M. The emerging role of long non-coding RNAs in endometrial cancer. Cancer Genet. 2016; 209:445–55.
7. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–8.
8. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011; 43:904–14.
9. Li T, Mo X, Fu L, Xiao B, Guo J. Molecular mechanisms of long noncoding RNAs on gastric cancer. Oncotarget. 2016; 7:8601–12.
10. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009; 136:629–41.
11. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nature Reviews Genetics. 2009; 10:155–9.
12. Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, Lin L, Yao H, Su F, Li D, Zeng M, Song E. A cytoplasmic NF-kappaB interacting long noncoding RNA blocks IkappaB phosphorylation and suppresses breast cancer metastasis. Cancer Cell. 2015; 27:370–81.
13. Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K. Aberrant

silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature. 2010; 465:175–81.

14. Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell. 2005; 122:553–63.
15. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19:92–105.
16. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010; 38:5366–83.
17. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. Nature. 2013; 495:384–8.
18. Yu G, Yao W, Gumireddy K, Li A, Wang J, Xiao W, Chen K, Xiao H, Li H, Tang K, Ye Z, Huang Q, Xu H. Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. Mol Cancer Ther. 2014; 13:3086–97.
19. Feng L, Tong R, Liu X, Zhang K, Wang G, Zhang L, An N, Cheng S. A network-based method for identifying prognostic gene modules in lung squamous carcinoma. Oncotarget. 2016; 7:18006–20.
20. Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol. 2011; 21:354–61.
21. Liu Q, Huang J, Zhou N, Zhang Z, Zhang A, Lu Z, Wu F, Mo YY. LncRNA loc285194 is a p53-regulated tumor suppressor. Nucleic Acids Res. 2013; 41:4976–87.
22. Zheng L, Li X, Gu Y, Lv X, Xi T. The 3'UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. Breast Cancer Res Treat. 2015; 150:105–18.
23. Ren K, Li Y, Lu H, Li Z, Wu K, Han X. Long Noncoding RNA HOTAIR Controls Cell Cycle by Functioning as a Competing Endogenous RNA in Esophageal Squamous Cell Carcinoma. Transl Oncol. 2016; 9:489–97.
24. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009; 9:153–66.
25. Hussain SP, Harris CC. Inflammation and cancer: an ancient link with novel potentials. Int J Cancer. 2007; 121:2373–80.
26. Ridolfi L, Petrini M, Fiammenghi L, Riccobon A, Ridolfi R. Human embryo immune escape mechanisms rediscovered by the tumor. Immunobiology. 2009; 214:61–76.
27. Yang D, Sun Y, Hu L, Zheng H, Ji P, Pecot CV, Zhao Y, Reynolds S, Cheng H, Rupaimoole R, Cogdell D, Nykter M, Broaddus R, et al. Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. Cancer Cell. 2013; 23:186–99.
28. An N, Shi X, Zhang Y, Lv N, Feng L, Di X, Han N, Wang G, Cheng S, Zhang K. Discovery of a Novel Immune Gene Signature with Profound Prognostic Value in Colorectal Cancer: A Model of Cooperativity Disorientation Created in the Process from Development to Cancer. PLoS ONE. 2015; 10: e0137171.
29. Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. Bioinformatics. 2010; 26:1057–63.
30. Wang XJ, Gulbahce N, Yu HY. Network-based methods for human disease gene prediction. Briefings in Functional Genomics. 2011; 10:280–93.
31. Zhang CL, Li CQ, Li J, Han JW, Shang DS, Zhang YP, Zhang W, Yao QL, Han L, Xu YJ, Yan W, Bao ZS, You G, et al. Identification of miRNA-Mediated Core Gene Module for Glioma Patient Prediction by Integrating High-Throughput miRNA, mRNA Expression and Pathway Structure. PLoS ONE. 2014; 9.
32. Song L, Rawal B, Nemeth JA, Haura EB. JAK1 activates STAT3 activity in non-small-cell lung cancer cells and IL-6 neutralizing antibodies can suppress JAK1-STAT3 signaling. Mol Cancer Ther. 2011; 10:481–94.
33. Gong L, Song J, Lin X, Wei F, Zhang C, Wang Z, Zhu J, Wu S, Chen Y, Liang J, Fu X, Lu J, Zhou C, et al. Serine-arginine protein kinase 1 promotes a cancer stem cell-like phenotype through activation of Wnt/beta-catenin signalling in NSCLC. J Pathol. 2016; 240:184–96.
34. Lin F, Chengyao X, Qingchang L, Qianze D, Enhua W, Yan W. CRKL promotes lung cancer cell invasion through ERK-MMP9 pathway. Mol Carcinog. 2015; 54:E35-44.
35. Pesch B, Kendzia B, Gustavsson P, Jockel KH, Johnen G, Pohlabeln H, Olsson A, Ahrens W, Gross IM, Bruske I, Wichmann HE, Merletti F, Richiardi L, et al. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case-control studies. Int J Cancer. 2012; 131:1210–9.
36. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014; 42: D92-7.
37. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34: D140-4.
38. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007; 27:91–105.
39. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. Nat Genet. 2005; 37:495–500.
40. Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet. 2008; 82:949–58.
41. Goh XY, Rees JR, Paterson AL, Chin SF, Marioni JC, Save V, O'Donovan M, Eijk PP, Alderson D, Ylstra B, Caldas C, Fitzgerald RC. Integrative analysis of array-

comparative genomic hybridisation and matched gene expression profiling data reveals novel genes with prognostic significance in oesophageal adenocarcinoma. Gut. 2011; 60:1317–26.

42. Walker N, Yenokyan G, Friberg IK, Bryce J. Patterns in coverage of maternal, newborn, and child health interventions: projections of neonatal and under-5 mortality to 2035. Lancet. 2013; 382:1029–38.

43. Delgado O, Batten KG, Richardson JA, Xie XJ, Gazdar AF, Kaisani AA, Girard L, Behrens C, Suraokar M, Fasciani G, Wright WE, Story MD, Wistuba, II, et al. Radiation-enhanced lung cancer progression in a transgenic mouse model of lung cancer is predictive of outcomes in human lung and breast cancer. Clin Cancer Res. 2014; 20:1610–22.