Integrated analysis reveals five potential ceRNA biomarkers in human lung adenocarcinoma

Yu Liu, Deyao Xie, Zhifeng He and Liangcheng Zheng

Department of Thoracic Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

ABSTRACT

Background: Competing endogenous RNAs (ceRNAs) are a newly identified type of regulatory RNA. Accumulating evidence suggests that ceRNAs play an important role in the pathogenesis of diseases such as cancer. Thus, ceRNA dysregulation may represent an important molecular mechanism underlying cancer progression and poor prognosis. In this study, we aimed to identify ceRNAs that may serve as potential biomarkers for early diagnosis of lung adenocarcinoma (LUAD).

Methods: We performed differential gene expression analysis on TCGA-LUAD datasets to identify differentially expressed (DE) mRNAs, lncRNAs, and miRNAs at different tumor stages. Based on the ceRNA hypothesis and considering the synergistic or feedback regulation of ceRNAs, a lncRNA–miRNA–mRNA network was constructed. Functional analysis was performed using gene ontology term and KEGG pathway enrichment analysis and KOBAS 2.0 software. Transcription factor (TF) analysis was carried out to identify direct targets of the TFs associated with LUAD prognosis. Identified DE genes were validated using gene expression omnibus (GEO) datasets.

Results: Based on analysis of TCGA-LUAD datasets, we obtained 2,610 DE mRNAs, 915 lncRNAs, and 125 miRNAs that were common to different tumor stages ($|\log_2(\text{Fold change})| \geq 1$, false discovery rate < 0.01), respectively. Functional analysis showed that the aberrantly expressed mRNAs were closely related to tumor development. Survival analyses of the constructed ceRNA network modules demonstrated that five of them exhibit prognostic significance. The five ceRNA interaction modules contained one lncRNA (FENDRR), three mRNAs (EPAS1, FOXF1, and EDNRB), and four miRNAs (hsa-miR-148a, hsa-miR-195, hsa-miR-196b, and hsa-miR-301b). The aberrant expression of one lncRNA and three mRNAs was verified in the LUAD GEO dataset. Transcription factor analysis demonstrated that EPAS1 directly targeted 13 DE mRNAs.

Conclusion: Our observations indicate that lncRNA-related ceRNAs and TFs play an important role in LUAD. The present study provides novel insights into the molecular mechanisms underlying LUAD pathogenesis. Furthermore, our study facilitates the identification of potential biomarkers for the early diagnosis and prognosis of LUAD and therapeutic targets for its treatment.

INTRODUCTION

Lung cancer, which is one of the most common malignancies worldwide, poses a serious threat to human health and life (*Jemal et al., 2011*; *Torre et al., 2015*). In recent years, there has been an increase in the morbidity of lung cancer owing to deterioration of the environment (*Cui et al., 2017*; *David et al., 2016*; *Haugen et al., 2000*; *Jin et al., 2014*; *Kooperstein, Schifrin & Leahy, 1965*; *Weiss, 1983*). Furthermore, lung cancer is difficult to diagnose early and associated with poor prognosis and low survival rate (*Bernaudin, 2010*; *Hardavella, George & Sethi, 2016*; *Onganer, Seckl & Djamgoz, 2005*; *Prim et al., 2010*; *Reinmuth et al., 2014*; *Tamura et al., 2015*). The prognosis of patients with lung cancer is not only related to its biological factors and immune status, but is also closely related with the therapeutic factors (*Fan, Zhang & Liu, 2014*; *Prim et al., 2010*; *Ringbaek et al., 1999*; *Truong, Viswanathan & Erasmus, 2011*); association between multiple factors leads to complex changes in the disease and makes prognosis difficult. Tumor staging is an important factor in determining prognosis, including survival in cancer patients (*Gong et al., 2017*; *Li et al., 2017*). Studies have shown that patients with late stage lung cancer display extremely poor prognosis, with invasion and metastasis, and additionally respond poorly to clinical treatment (*Mattern, Koomagi & Volm, 2002*; *Prim et al., 2010*). Therefore, early detection of cancer should enable administration of effective treatment and considerably improve patient prognosis. Although the prognosis of non-small cell lung cancer (NSCLC) has been greatly improved following the introduction of multidisciplinary treatment, the prognosis of lung cancer patients and treatment efficiency remain poor (*Fan, Zhang & Liu, 2014*; *Huang et al., 2012*; *Miura et al., 1996*; *Pujol et al., 1994*); therefore, the identification of biological markers and therapeutic targets for the early detection and treatment of recurrent lung tumor is critical.

Competing endogenous RNAs (ceRNAs) are a group of regulatory RNA molecules that compete with other RNA molecules to bind specific miRNAs, thereby regulating target gene expression (*Phelps et al., 2016*; *Shukla, Singh & Barik, 2011*; *Tay et al., 2011*). Many studies have suggested that miRNA-mediated ceRNA regulatory mechanisms play crucial roles in tumor occurrence and development (*Bartel, 2009*; *Hansen et al., 2013*; *Tay, Rinn & Pandolfi, 2014*). Therefore, the identification of ceRNA biomarkers related to lung cancer should enable the development of effective strategies for the diagnosis and treatment of lung cancer.

In this study, we aimed to identify differentially expressed genes (DEGs) associated with lung adenocarcinoma (LUAD) development and construct a ceRNA regulatory network using the TCGA-LUAD dataset. Further, we examined the relationship of the identified genes and ceRNA interaction modules with overall survival and prognosis. Through our study, we sought to gain new insights into the molecular mechanisms underlying LUAD and identify potential biomarkers for early diagnosis of this disease.

MATERIALS AND METHODS

Data source and preprocessing

Lung adenocarcinoma RNA-seq and lncRNA gene expression level three data were obtained from the database Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/docs/publications/tcga/). After an initial filter (excluding samples with uncertain tumor stage),

569 samples were included for further study. Patient clinical data, including outcome and staging information, were also downloaded. Of the 569 samples, 510 were LUAD samples (168 with T1 stage, 276 with T2 stage, 47 with T3 stage, and 19 with T4 stage) and 59 were normal samples.

Screening of genetically altered genes

Gene raw read counts were used to perform differential expression analysis with DESeq2 (v 1.18.1) (*Love, Huber & Anders, 2014*), a R package that uses a model based on the negative binomial distribution and is widely used for RNA-seq data differential analysis. Differentially expressed genes were defined by two criteria: false discovery rate (FDR) < 0.01 and fold change (FC) $\geq$ 2. In order to filter out the low-quality DEGs, we performed the following additional filter strategy. For mRNA and miRNA, we only retained DEGs that showed FPKM or RPKM larger than the threshold of one for at least 10% samples; for lncRNA DEGs, the threshold was set to 0.1. We then compared the normal tissue sample with different tumor stage samples (T1, T2, T3, and T4) to identify common DEGs for further analysis.

Construction of lncRNA-associated ceRNA network (gene-set construction)

Based on the relationship between lncRNA, mRNA, miRNA, a ceRNA network was constructed using the following steps: 1) Both up-regulated and down-regulated LUAD-specific RNAs were chosen to construct the ceRNA network; 2) miRanda software was used to predict the miRNA–lncRNA interactions, and interactions in starBase (*Li et al., 2014*) were also included; 3) miRNA–mRNA interactions in miRTarbase and starBase were used; 4) Correlation between LUAD-specific lncRNA and mRNA was calculated with WGCNA (*Zhang & Horvath, 2005*) corAndPvalue function and correlation pairs with coefficient > 0.8 and *p*-value < 0.05 were retained; 5) The hypergeometric method was used to test the enrichment significance of miRNAs between mRNA and lncRNA. The enrichment significance was calculated by using the following formula:

$$P = \sum_{i=c}^{\min(K,\, n)} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

In the above formula, N represents the total number of miRNAs, K is the number of miRNAs targeting mRNA, n is the number of miRNAs targeting lncRNA, and c is the number of miRNAs shared by the mRNA and lncRNA. Only ceRNA modules with FDR < 0.05 were considered.

Gene functional enrichment analysis

In order to investigate key mRNAs at molecular and functional level, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway function enrichment analysis was performed. GOseq, which uses Wallenius' noncentral hypergeometric

distribution model taking gene length bias into account, was used to perform GO enrichment analysis. Gene ontology terms with p-value less than 0.05 were considered significantly enriched. KOBAS 2.0 software was used to test statistical enrichment in KEGG pathways, and those with a Fisher's exact test p-value of less than 0.05 was considered to be significantly enriched (*Xie et al., 2011*).

Transcription factor (TF) analysis

Co-expression networks of LUAD-mRNA expression data sets were generated using Genie3 with EPAS1 and FOXF1 as Transcription factors (TFs) of interest. A threshold of 0.005 was used to filter the co-expression network. Additionally, Spearman correlations between the TFs and LUAD-specific mRNAs were calculated, and TF-mRNA pairs with coefficient > 0.03 were retained as this value represents positive correlation. Using this procedure, we obtained a TF co-expression draft regulatory network. As the analysis was only based on co-expression, the results may include numerous indirect targets (such as downstream effects). To identify genes that are most likely the direct targets, we used RcisTarget (*Aibar et al., 2017*; *Imrichova et al., 2015*) to perform cis-regulatory motif analysis of each TF regulon. RcisTarget is an R package that identifies TF binding motifs that are over-represented in a gene list. We used the database, scoring 500 bp upstream of TSS, and only motifs with a Normalized Enrichment Score (NES) > 3.0 were considered associated with TFs of interest.

Validation of gene expression with gene expression omnibus data

In order to validate the key differentially expressed (DE) molecules, we examined four datasets from among the gene expression omnibus (GEO) datasets. Of the selected datasets, the GSE10072 study contained 58 LUAD tumor and 49 non-tumor tissue mRNA expression data, GSE32863 contained 58 LUAD tumor and 58 non-tumor tissue mRNA expression data, GSE85716 contained six LUAD tumor and six non-tumor tissue lncRNA expression data, and GSE104854 contained three LUAD (pooled from nine tissues) and three non-tumor (pooled from nine tissues) expression data. GSE10072, GSE32863, and GSE85716 datasets, which were derived from microarray data, were analyzed with GEO2R online. The GSE104854 dataset, which was obtained by high-throughput sequencing, was analyzed with the DESeq2 R package based on read counts.

Statistical analysis

To investigate the impact of the expression level of RNAs on the prognostic survival of patients, survival analysis was performed. In the analysis, Kaplan–Meier survival analyses and log-rank test were performed to study the relationship between RNA expression states (cutoff point: median value) and survival time. Univariate cox proportional hazards regression was applied to identify the RNAs associated with overall survival. In order to evaluate the potential of dysregulated ceRNA modules as biomarkers, the risk scoring classifier was constructed. For each ceRNA module, we calculated the risk score for multivariate survival analysis to determine the prognostic influence of the RNAs as a whole. All statistical analyses were performed using R version 3.4.3 software (*R Core Team, 2017*).
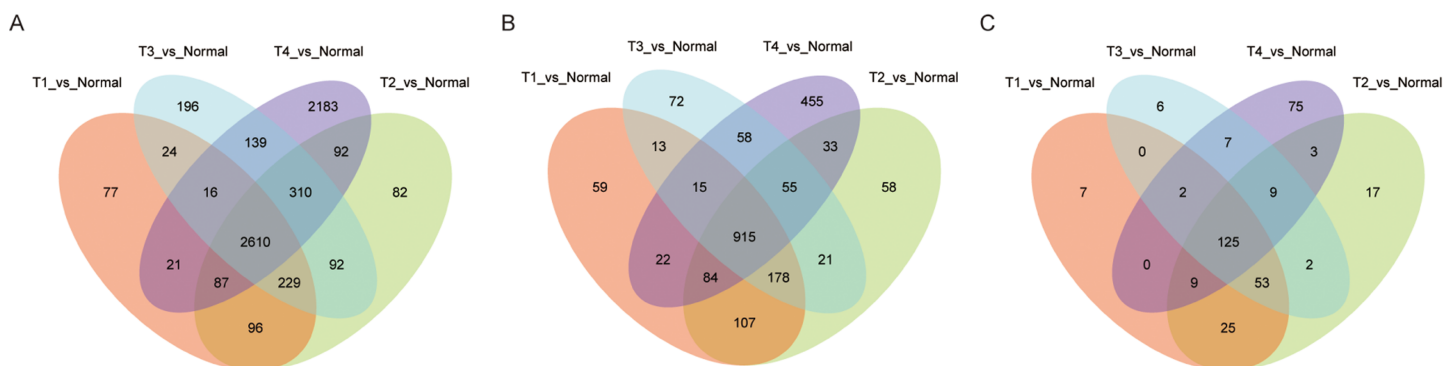
Liu et al. (2019), *PeerJ*, DOI 10.7717/peerj.6694

4/17

Figure 1 Significantly expressed genes in LUAD. (A) The Venn diagram of mRNA DEGs, the circle represents mRNA DEGs between normal tissue and tumor stage tissue; (B) the Venn diagram of lncRNA DEGs; (C) the Venn diagram of miRNA DEGs.

RESULTS

Differentially expressed RNAs in LUAD

Among the LUAD datasets, expression data for 19,669 mRNAs, 7,309 lncRNAs, and 1,882 miRNAs were extracted from TCGA and analyzed with R package (DESeq2). In this analysis, we compared adjacent normal lung tissue with lung cancer subtypes defined by pathology stage (T1-T4), respectively. When we combined these four groups and analyzed for DE RNAs, 2,160 mRNAs (1,527 up- and 1,083 down-regulated), 915 lncRNAs (662 up- and 253 down-regulated), and 125 miRNAs (73 up- and 52 down-regulated) showed consistently differential expression (FC $\geq$ 2, FDR < 0.01) (Figs. 1A–1C). Figs. 2A–2C shows the heatmap of the common DEGs; the lung tumor tissue samples could be easily distinguished from the adjacent non-tumor lung tissues from the heat map. Based on these data, DE mRNAs, lncRNAs, and miRNAs were selected for further analysis.

Functional enrichment of mRNA DEGs

In order to understand the functions of aberrantly expressed genes in this study, a total of 2,610 aberrantly expressed genes were analyzed. We used GOseq for GO annotation and KOBAS 2.0, which includes the KEGG database, to classify and analyze the potential gene functions in the pathways. Our analysis revealed an enrichment of 2,879 GO terms and 30 KEGG pathways (*p* value < 0.05) (Table S1). Gene ontology analyses were classified into three functional groups: molecular function group, biological process group, and cellular component group (Fig. 3). Enrichment of these DEGs represents a measure of significance of a function. As shown in Fig. 3, the most enriched GO biological process terms were "single-organism cellular process" and "single-multicellular organism process." Other significant GO terms included "cell proliferation," "cell adhesion," "system development," "cell migration," and "cell adhesion." The significant pathways identified by KEGG pathway analysis included viral carcinogenesis, systemic lupus erythematosus, cell adhesion molecules, and p53 signaling pathway (Fig. 4). Our analysis indicated that 2,610 dysregulated mRNAs were involved in signaling pathways related to environmental information processing and human diseases such as cancer.
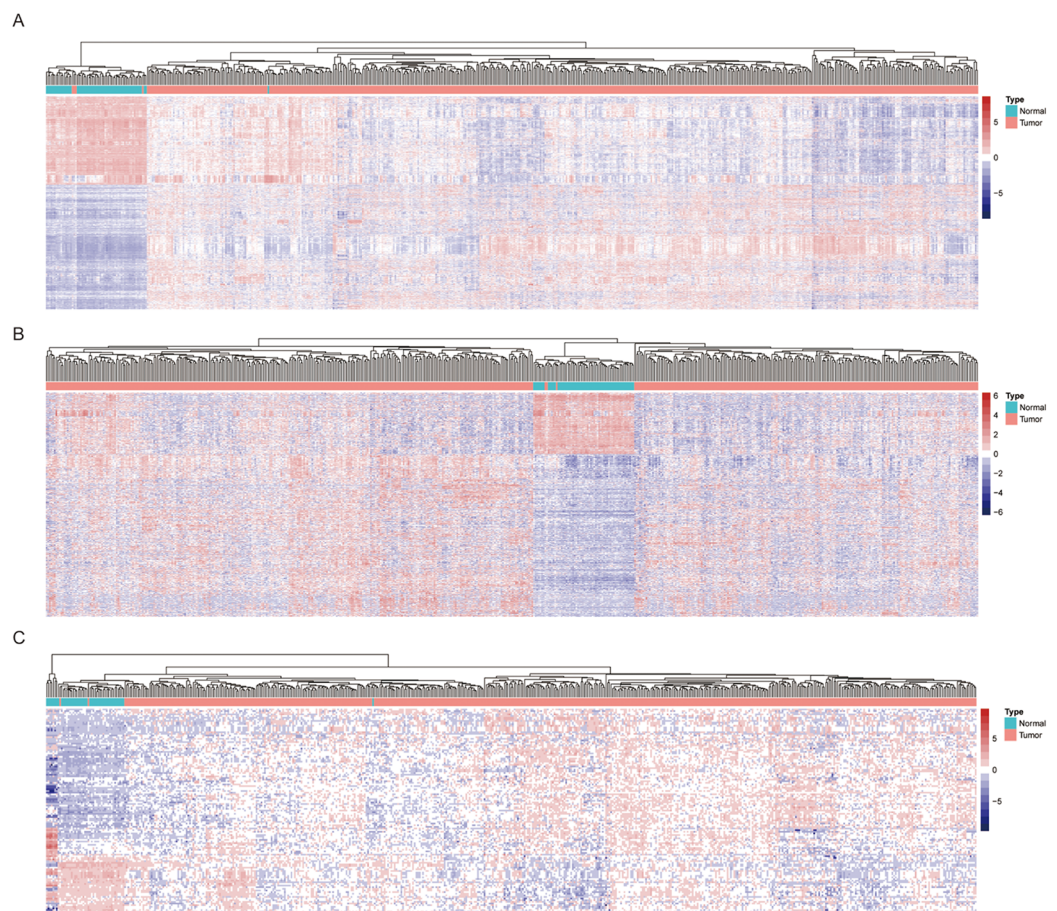
Figure 2 Heatmap of significantly expressed common genes between different pathology stages of LUAD. (A) mRNA DEGs; (B) lncRNA DEGs; (C) miRNA DEGs.

Full-size ⬜ DOI: 10.7717/peerj.6694/fig-2

Competing endogenous RNA network analysis and integrated ceRNA network construction

Based on the miRNAs and mRNAs described in Fig. 2, starBase v2.0 database and miRanda were used to predict miRNA-targeted RNAs. Then, combing miRNA–lncRNA interactions with the miRNA–mRNA interactions, an integrated lncRNA–miRNA–mRNA ceRNA network was established. As shown in Fig. 5, the network contained 107 interactions, and included 5 lncRNAs, 20 mRNAs, and 34 miRNAs (Table S2). After ranking the molecules according to their degree, FENDRR, which is a lncRNA, showed the highest degree (degree = 38). *FOXF1*, *EDNRB*, and *EPAS1* are all mRNA molecules, ranking the second, third, and fourth respectively. In other studies, *FENDRR* was demonstrated to be down-regulated in breast and gastric cancers and to participate in regulating tumor malignancy (*Li et al., 2018*; *Xu et al., 2014*). FOXF1 and EPAS1 are TFs that have been demonstrated to participate in tumor progression (*Putra et al., 2015*; *Tamura et al., 2014*). EDNRB is a G protein-coupled receptor that activates a phosphatidylinositol-calcium second messenger system, and is also closely related with cancer (*Chen et al., 2013*; *Wuttig et al., 2012*).
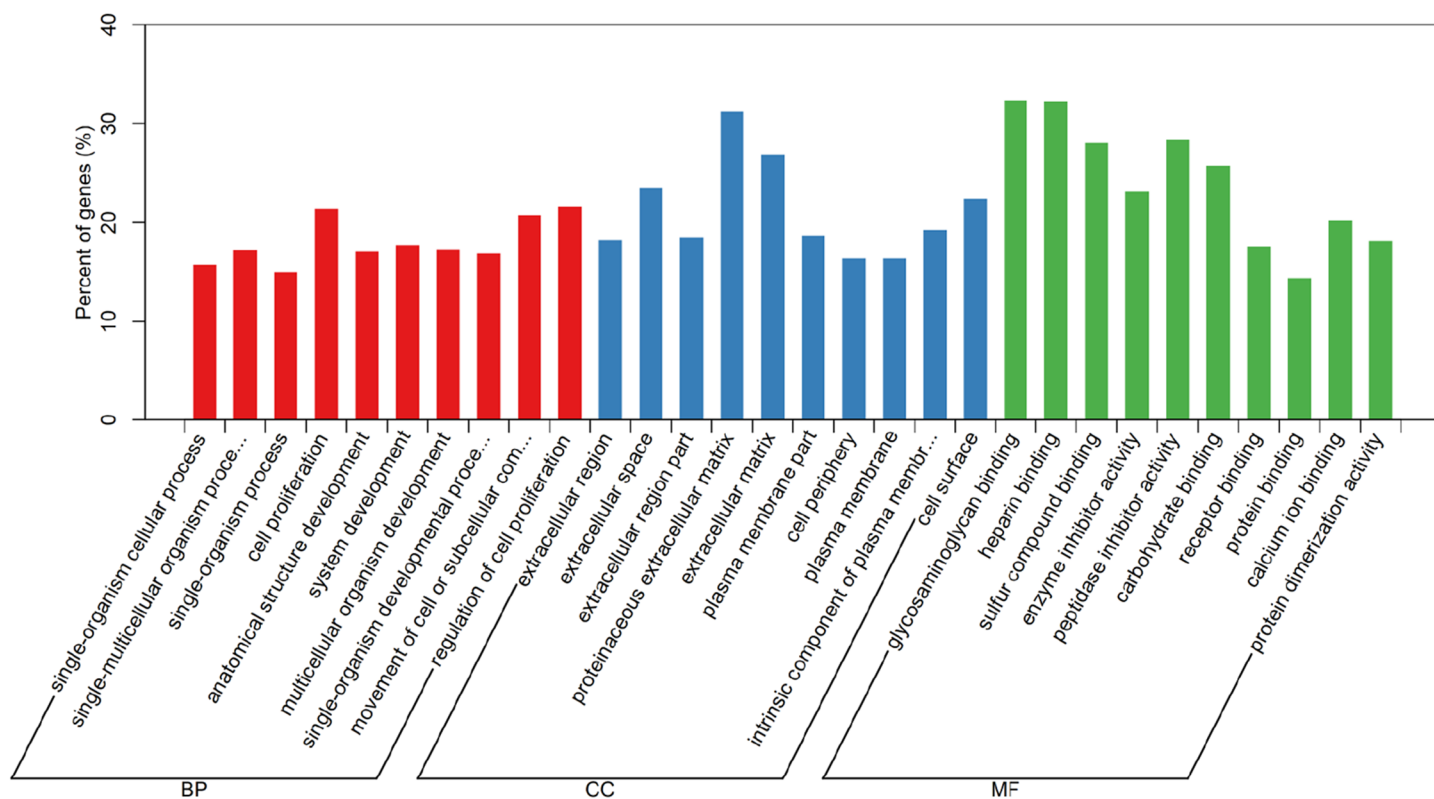
Figure 3 Gene Ontology analysis and significant enriched GO terms of DEGs in LUAD. GO analysis classified the DEGs into three groups (i.e., molecular function, biological process and cellular component). Full-size ◲ DOI: 10.7717/peerj.6694/fig-3

Key ceRNAs and their association with clinical features

In order to further explore the key role of ceRNAs in the occurrence and development of LUAD, we aimed to identify prognostic-specific ceRNAs. We calculated expression of each ceRNA and performed prognosis analyses (e.g. of overall survival time) to determine the survival-significant lncRNA–miRNA–mRNA interactions based on Cox analysis (p value < 0.05). As presented in Fig. 6 and Table 1, univariate Cox regression analysis showed that EDNRB-FENDRR-hsa-mir-196b, EPAS1-FENDRR-hsa-mir-148a, FOXF1-FENDRR-hsa-mir-148a, FOXF1-FENDRR-hsa-mir-195, and FOXF1-FENDRR-hsa-mir-301b ceRNA interactions were associated with the overall survival of patients with LUAD. Kaplan–Meier survival curves indicated that EDNRB-FENDRR-hsa-mir-196b interaction was negatively correlated with overall survival, whereas EPAS1-FENDRR-hsa-mir-148a, FOXF1-FENDRR-hsa-mir-148a, FOXF1-FENDRR-hsa-mir-195, and FOXF1-FENDRR-hsa-mir-301b were positively correlated with survival.

Transcription factor cis-target analysis

Transcription factors play important roles in the regulation of biological processes by binding enhancer or promoter regions of DNA, thereby activating or suppressing gene expression. This allows genes to be expressed in the right cell at the appropriate time. In this study, we found that EPAS1 and FOXF1 showed a high degree of representation
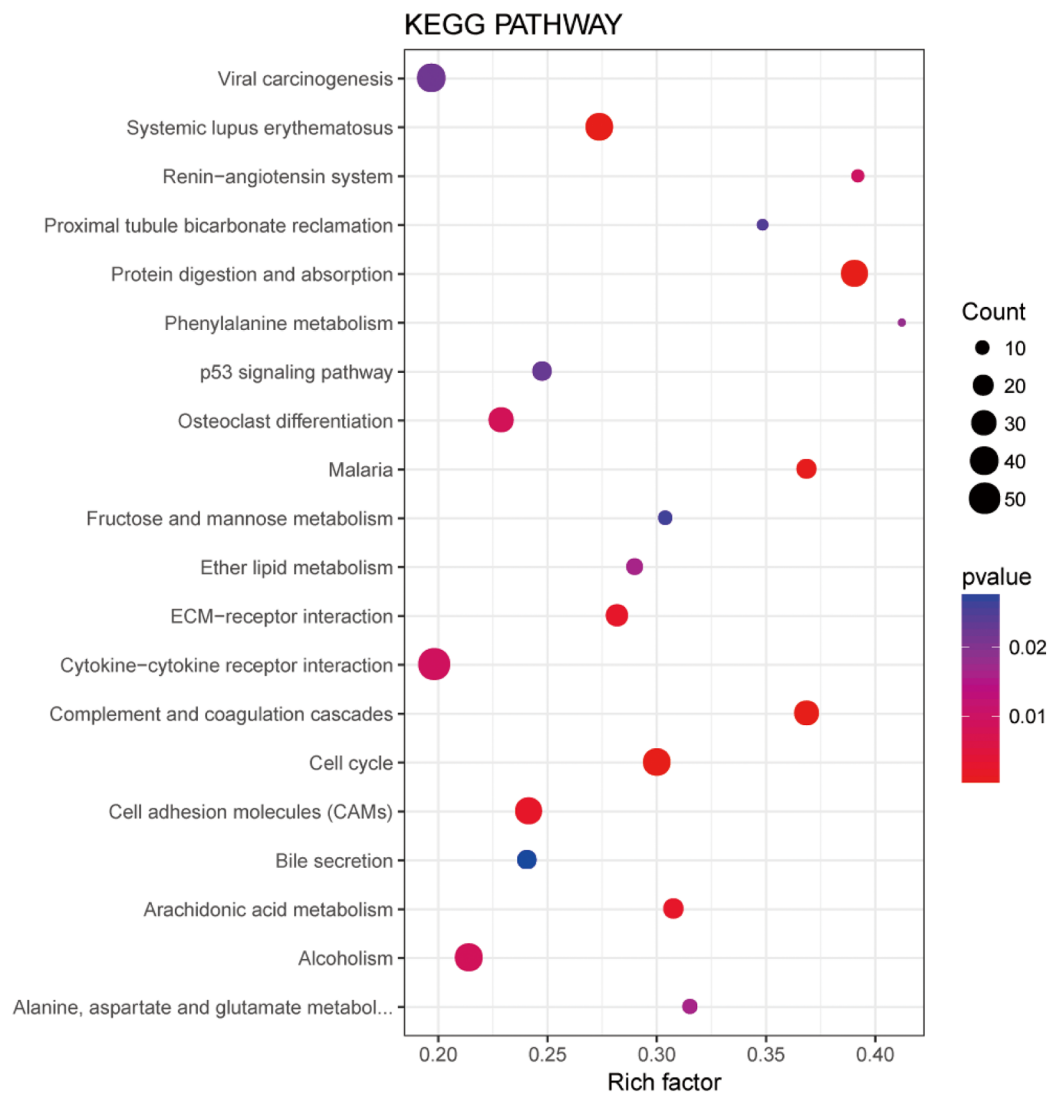
Figure 4 Significantly enriched pathway terms in LUADs and 5ST. The functional and signaling pathway enrichment were conducted using KOBAS 2.0 software.

Full-size 🖼 DOI: 10.7717/peerj.6694/fig-4

in the ceRNA network. This finding suggests that these two TFs play an important role in LUAD and may be regulated by a ceRNA-based mechanism. In order to verify the direct targets of the two TFs, we analyzed them further. Co-expressed genes may be regulated by the same or similar transcription factors; however, they may include numerous indirect targets. Therefore, we performed binding motif enrichment analysis to identify direct target genes of the two TFs. In this study, we use GENIE3 to perform TF-genes co-expression analysis and RcisTarget to analyze enriched motifs in the TSS upstream of the gene-sets. As shown in Fig. 7, the *EPAS1* co-expressed gene-set was enriched for the "cisbp__M6212" motif (NES = 3.21); likely direct gene targets of EPAS1 include *ADRB2, CALCRL, EMP2, FHL1, GRIA1, LGI3, LIMS2, NCKAP5, RXFP1, SLC6A4, SMAD6, STX11*, and *TMEM100*. However, the *FOXF1* co-expressed gene-set did not show enrichment for any motifs.
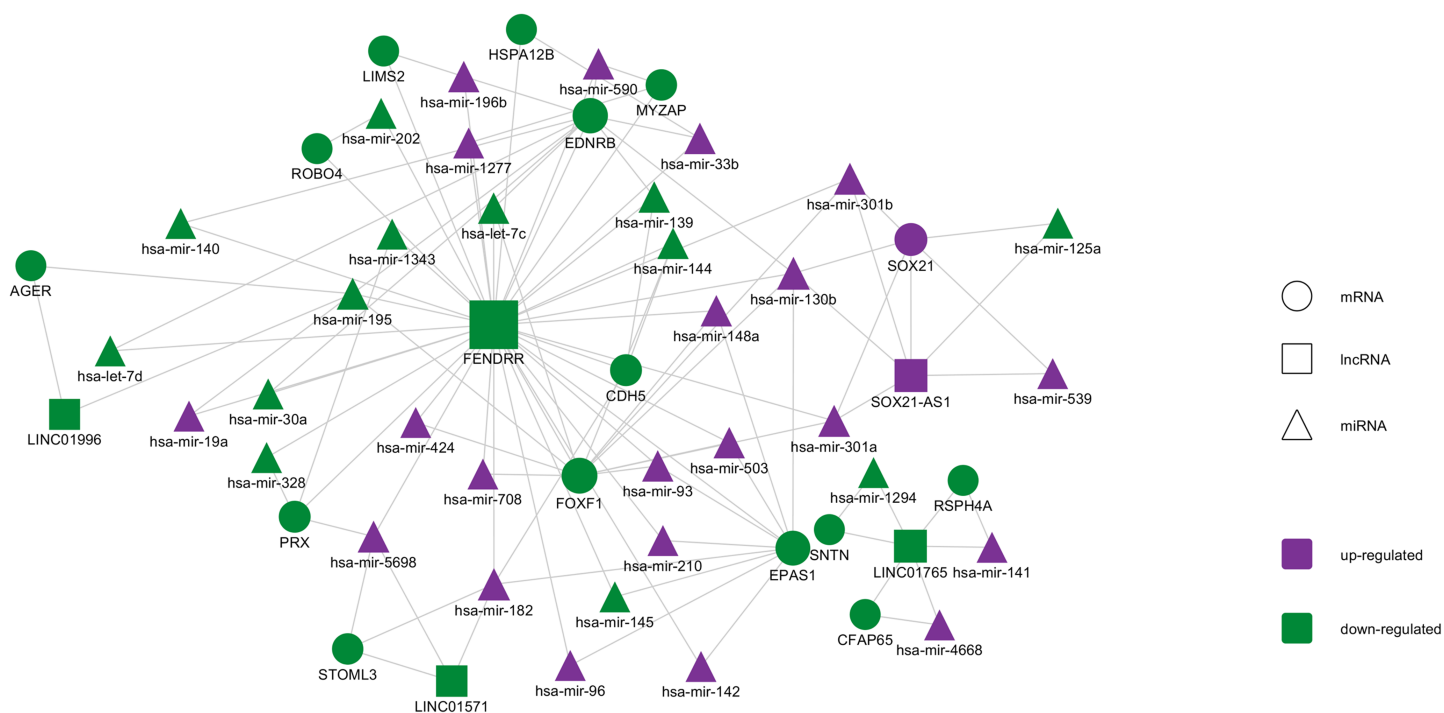
Figure 5 The lncRNA–miRNA–mRNA ceRNA network constructed from DEGs. Purple represents up-regulated genes and yellow represents down-regulated genes. Circles represent mRNA, rectangles represent lncRNA, and triangles represent miRNA.

Validation of the expression of key RNA molecules with GEO data

Finally, four reported studies were screened out from the GEO to verify the differential expression of key mRNAs and lncRNAs in LUAD. GSE10072 and GSE32863 were used for *EPAS1*, *FOXF1*, and *EDNRB* mRNA verification. GSE85716 and GSE104854 were used for *FENDRR* differential expression. As shown in Fig. 8 and Table 2, all four molecules showed significantly lower levels of expression in tumor datasets, which is consistent with the results of our study.

DISCUSSION

Non-small cell lung cancer, which mainly includes adenocarcinoma and squamous cell carcinoma, accounts for about one-sixth of cancer deaths in the global population (*Jemal et al., 2011*; *Torre et al., 2015*). Despite numerous advances in the treatment of lung cancer, the prognosis of this disease remains poor, and the 5-year overall survival rate is only 11% (*Huang et al., 2012*; *Miura et al., 1996*). Therefore, elucidation of the underlying mechanisms is indispensable for diagnosing, preventing, and treating NSCLC.

With the development of high-throughput sequencing technology, accumulating omics data reveal that the occurrence and development of the disease is closely related to multiple factors, including genomic mutations, expression profiles, copy number variants, and structural variations (*Balbin et al., 2013*; *Kikutake & Yahara, 2016*; *Kim et al., 2015*; *Yan et al., 2017*). In recent years, several studies have found that multiple RNA molecules are closely related to tumor progression (*Chen et al., 2018*; *He et al., 2018*;
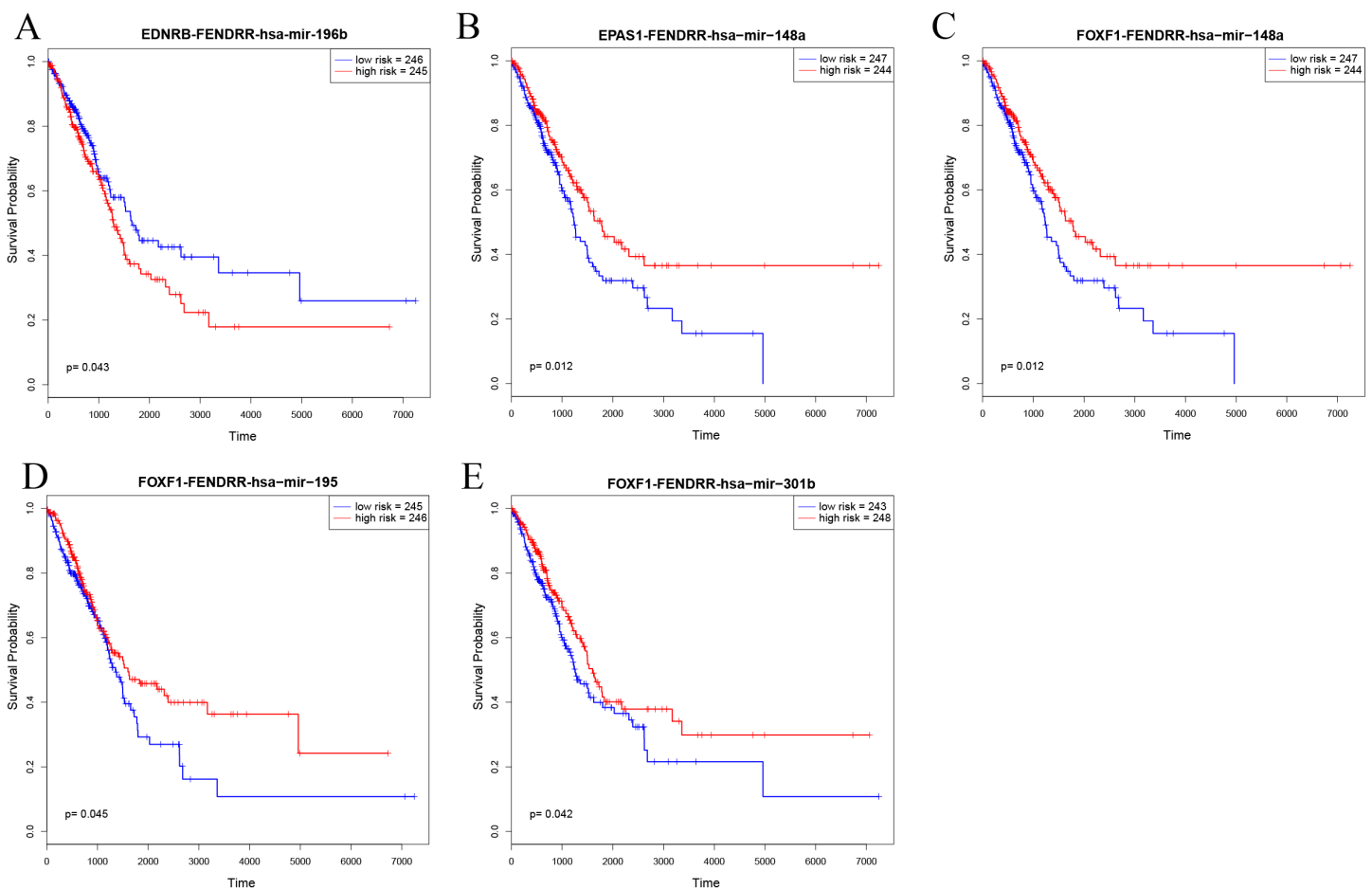
Figure 6 Kaplan–Meier survival curves for five ceRNA pairs associated with overall survival. (A) EDNRB-FENDRR-hsa-mir-196b. (B) EPAS1-FENDRR-hsa-mir-148a. (C) FOXF1-FENDRR-hsa-mir-148a. (D) FOXF1-FENDRR-hsa-mir-195. (E) FOXF1-FENDRR-hsa-mir-301b.

Full-size 🖼 DOI: 10.7717/peerj.6694/fig-6

Table 1 Gene fold change in five ceRNA pairs.

Gene	T1_vs_Normal		T2_vs_Normal		T3_vs_Normal		T3_vs_Normal	
	log2FC	FDR	log2FC	FDR	log2FC	FDR	log2FC	FDR
EPAS1	−2.64	5.65E-97	−2.76	1.1E-127	−2.49	1.3E-64	−2.61	9.45E-41
FOXF1	−2.44	6.68E-55	−2.96	2E-126	−2.58	1.2E-45	−2.63	8.73E-30
EDNRB	−3.27	2.09E-78	−3.75	9.2E-123	−3.38	3.4E-60	−2.94	8.47E-38
FENDRR	−3.48	2.85E-54	−4.37	2.7E-112	−3.78	1.62E-42	−3.89	1.94E-42
hsa-mir-148a	1.51	1.31E-24	1.46	1.51E-23	1.45	4.7E-11	1.51	1.07E-07
hsa-mir-195	−1.98	6.47E-37	−2.25	5.03E-68	−2.28	2.9E-30	−2.08	6.23E-13
hsa-mir-196b	3.00	1.59E-21	4.42	7.94E-35	2.98	5.6E-15	4.21	1.08E-20
hsa-mir-301b	3.48	2.62E-22	4.14	7.95E-28	2.94	9.5E-15	3.02	1.82E-05

Men, Liu & Ren, 2018; *Shang et al., 2017*; *Song et al., 2018*; *Tang et al., 2017*; *Wang et al., 2018*; *Yu et al., 2017*). For example, lncRNA-DANCR enhances the stemness feature of cancer cells to increase the degree of malignancy of lung cancer (*Lu et al., 2018*;
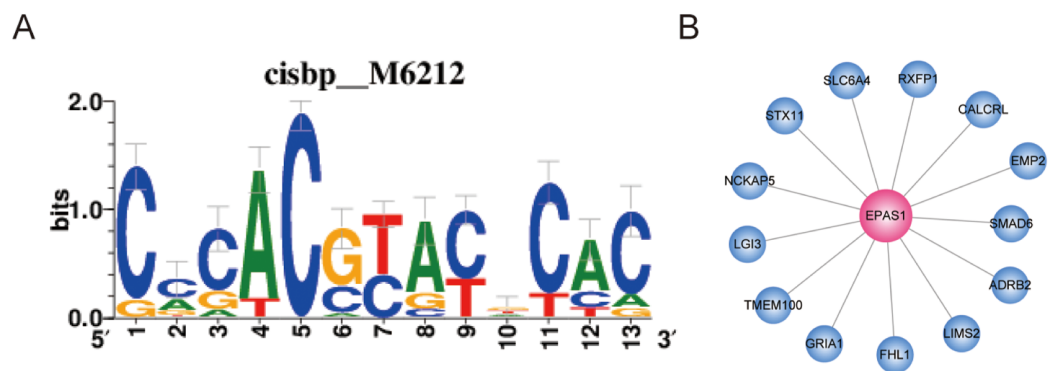
Figure 7 Co-expressed genes analysis for EPAS1. (A) EPAS1 co-expressed gene-set was enriched for motif "cisbp__M6212." (B) EPAS1-regulated genes in LUAD.

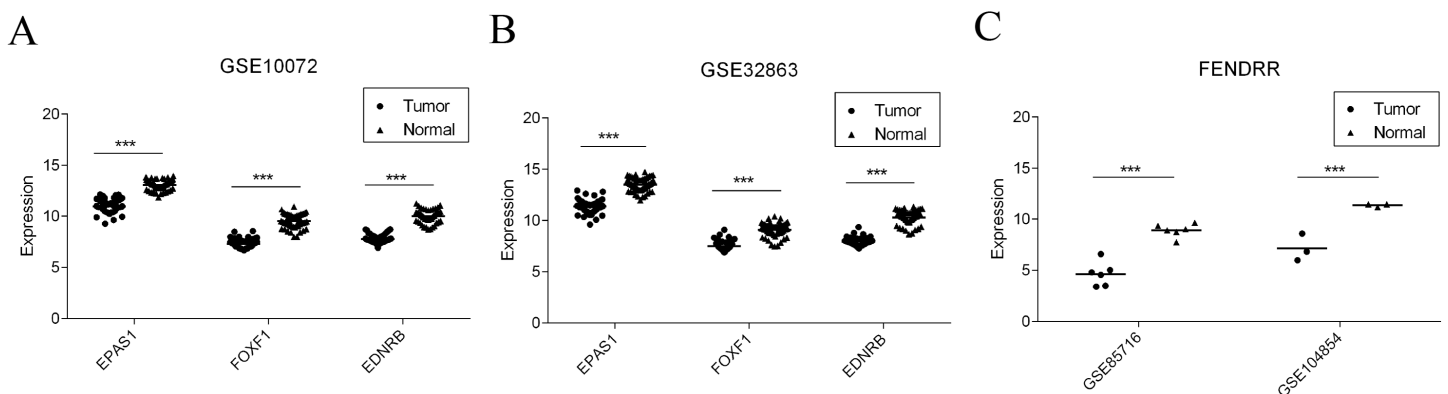Full-size 🖼 DOI: 10.7717/peerj.6694/fig-7

Figure 8 Validation of gene expression in GEO dataset. (A) EPAS1, FOXF1, and EDNRB expression level in GSE10072. (B) EPAS1, FOXF1 and EDNRB expression level in GSE32863. (C) FENDRR expression level in GSE85716 and GSE104854.

Full-size 🖼 DOI: 10.7717/peerj.6694/fig-8

Table 2 EPAS1, FOXF1, EDNRB and FENDRR expression in GEO dataset.

Study	mRNA						lncRNA	
	EPAS1		FOXF1		EDNRB		FENDRR	
	log2FC	FDR	log2FC	FDR	log2FC	FDR	log2FC	FDR
GSE10072	−2.087	3.08E-33	−2.228	2.17E-35	−2.299	4.30E-35	–	–
GSE32863	−2.152	3.05E-33	−1.616	2.69E-28	−2.385	2.35E-44	–	–
GSE85716	–	–	–	–	–	–	−4.25009	0.000343
GSE104854	–	–	–	–	–	–	−4.12757	3.96E-08

Zhen et al., 2018); miRNA-221 was reported to regulate the process of tumor differentiation, proliferation, migration, and invasion (*Lv et al., 2014*; *Yin, Xu & Li, 2017*). These RNA molecules represent biological markers for the prognosis of lung cancer. In this study, we found that multiple novel ceRNA molecules can be used as novel prognostic markers; further, we examined the putative molecular mechanisms underlying the development of tumors.

miRNAs modulate the expression of target genes by regulating the transcription and stability of their mRNAs or lncRNAs. Investigation of the lncRNA-mRNA co-expression network is important for analysis of the function and regulatory mechanisms of lncRNAs (*Fang, Wang & Li, 2018*). In this study, analysis showed that a number of mRNAs, lncRNAs, and miRNAs that were DE were common at each stage of lung cancer. Gene ontology and KEGG analysis showed that the DEGs were mainly involved in "single-organism cellular process," "single-multicellular organism process," "cell proliferation," "cell adhesion," "system development," "cell migration," and "cell adhesion biological processes." Further analysis revealed that the up-regulated genes were involved in the mitotic cell cycle and nuclear division process (Table S3), while down-regulated genes were involved in vasculature development, cell motility, cell migration, and cell proliferation (Table S4). Kyoto encyclopedia of genes and genomes analysis found that the up-regulated genes were mainly involved in cancer and other related signaling pathways such as the p53 signaling pathway.

Survival curve analysis reveals the correlation between gene expression and patient prognosis. Our study found that the five most important ceRNA-related RNA molecules were significantly correlated with the survival rate of patients with lung cancer ($p < 0.01$), and most of the genes were highly correlated with cancer. The five ceRNA interactions consisted of one lncRNA, three mRNAs, and four miRNAs. The lncRNA, FENDRR (*Grote et al., 2013*), which is produced from a spliced long non-coding RNA transcribed bidirectionally with FOXF1 on the opposite strand, is down-regulated in both breast cancer and gastric cancer. In breast cancer, lower expression level of FENDRR was associated with shorter overall survival and progression-free survival in breast cancer patients. FENDRR knockdown promoted breast cancer cell proliferation and migration, and suppressed cell apoptosis; in contrast, its overexpression inhibited tumor growth in a xenograft model (*Li et al., 2018*). FENDRR was also reported to be down-regulated in gastric cancer cell lines and tissues. Further, FENDRR expression was negatively correlated with tumor metastasis and stage (*Xu et al., 2014*). In our study, FENDRR was also down-regulated in LUAD, and showed the highest degree in ceRNA network, indicating its important role in LUAD. The three mRNAs, *FOXF1*, *EDNRB*, and *EPAS1*, ranked 2nd, 3rd, and 4th, respectively in the ceRNA network. FOXF1 and EPAS1 are both TFs that have been associated with tumor malignancy (*Putra et al., 2015*; *Tamura et al., 2014*). Our TF analysis indicated that EPAS1 directly regulates 13 molecules associated with LUAD, namely ADRB2, CALCRL, EMP2, FHL1, GRIA1, LGI3, LIMS2, NCKAP5, RXFP1, SLC6A4, SMAD6, STX11, and TMEM100. EDNRB, a G protein-coupled receptor, is also closely related with tumor (*Chen et al., 2013*; *Wuttig et al., 2012*). Our findings suggest that these molecules represent potential cancer-related biomarkers.

CONCLUSIONS

In summary, in the present study, we comprehensively analyzed the expression of key RNA molecules in the progression of lung cancer and identified key RNA molecules (including mRNA/lncRNA/miRNA) that were abnormally expressed in different stages of this disease. Further, using the above-mentioned key RNA molecules to construct a ceRNA–gene interaction network, it was found that several FOXF1- and EDNRB-related

ceRNA molecules play an important role in the development and progression of lung cancer, and can affect the prognosis of this disease. This series of related ceRNA molecules is expected to provide a novel basis for diagnosis and evaluation of prognosis, and represents potential new targets for the treatment of lung cancer.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Yu Liu conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Deyao Xie performed the experiments, analyzed the data, prepared figures and/or tables.
- Zhifeng He performed the experiments, analyzed the data, prepared figures and/or tables.
- Liangcheng Zheng conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability
The following information was supplied regarding data availability:
The origin RNAseq data used in our study were all downloaded from the TCGA data portal: https://portal.gdc.cancer.gov/;
GEO data link of mRNA: GSE10072; GSE32863;
GEO data link of lncRNA: GSE85716; GSE104854.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6694#supplemental-information.

REFERENCES

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14(11)**:1083–1086 DOI 10.1038/nmeth.4463.

Balbin OA, Prensner JR, Sahu A, Yocum A, Shankar S, Malik R, Fermin D, Dhanasekaran SM, Chandler B, Thomas D, Beer DG, Cao X, Nesvizhskii AI, Chinnaiyan AM. 2013. Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nature Communications* **4(1)**:2617 DOI 10.1038/ncomms3617.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136(2)**:215–233 DOI 10.1016/j.cell.2009.01.002.

Bernaudin JF. 2010. Molecular characteristics of lung cancer. *Bulletin du Cancer* **97(11)**:1323–1335 DOI 10.1684/bdc.2010.1211.

Chen J, Wang S, Jia S, Ding G, Jiang G, Cao L. 2018. Integrated analysis of long non-coding RNA and mRNA expression profile in pancreatic cancer derived exosomes treated dendritic cells by microarray analysis. *Journal of Cancer* **9(1)**:21–31 DOI 10.7150/jca.21749.

Chen C, Wang L, Liao Q, Huang Y, Ye H, Chen F, Xu L, Ye M, Duan S. 2013. Hypermethylation of EDNRB promoter contributes to the risk of colorectal cancer. *Diagnostic Pathology* **8(1)**:199 DOI 10.1186/1746-1596-8-199.

Cui J-J, Wang L-Y, Zhu T, Gong W-J, Zhou H-H, Liu Z-Q, Yin J-Y. 2017. Gene-gene and gene-environment interactions influence platinum-based chemotherapy response and toxicity in non-small cell lung cancer patients. *Scientific Reports* **7(1)**:5082 DOI 10.1038/s41598-017-05246-8.

David SP, Wang A, Kapphahn K, Hedlin H, Desai M, Henderson M, Yang L, Walsh KM, Schwartz AG, Wiencke JK, Spitz MR, Wenzlaff AS, Wrensch MR, Eaton CB, Furberg H, Mark Brown W, Goldstein BA, Assimes T, Tang H, Kooperberg CL, Quesenberry CP, Tindle H, Patel MI, Amos CI, Bergen AW, Swan GE, Stefanick ML. 2016. Gene by environment investigation of incident lung cancer risk in African-Americans. *EBioMedicine* **4**:153–161 DOI 10.1016/j.ebiom.2016.01.002.

Fan X, Zhang Z, Liu Y. 2014. Therapeutic efficacy and prognosis in the treatment of lung cancer by video-assisted thoracoscopic surgery. *Chinese Medical Journal* **127**:2096.

Fang L, Wang H, Li P. 2018. Systematic analysis reveals a lncRNA-mRNA co-expression network associated with platinum resistance in high-grade serous ovarian cancer. *Investigational New Drugs* **36(2)**:187–194 DOI 10.1007/s10637-017-0523-3.

Gong B, Li Y, Cheng Z, Wang P, Luo L, Huang H, Duan S, Liu F. 2017. GRIK3: A novel oncogenic protein related to tumor TNM stage, lymph node metastasis, and poor prognosis of GC. *Tumour Biology* **39(6)**:1010428317704364 DOI 10.1177/1010428317704364.

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, Herrmann BG. 2013. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental Cell* **24(2)**:206–214 DOI 10.1016/j.devcel.2012.12.012.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495(7441)**:384–388 DOI 10.1038/nature11993.

Hardavella G, George R, Sethi T. 2016. Lung cancer stem cells—characteristics, phenotype. *Translational Lung Cancer Research* **5(3)**:272–279 DOI 10.21037/tlcr.2016.02.01.

Haugen A, Ryberg D, Mollerup S, Zienolddiny S, Skaug V, Svendsrud DH. 2000.
Gene–environment interactions in human lung cancer. *Toxicology Letters* **112–113**:233–237
DOI 10.1016/S0378-4274(99)00275-1.

He ZH, Qin XH, Zhang XL, Yi JW, Han JY. 2018. Long noncoding RNA GIHCG is a potential
diagnostic and prognostic biomarker and therapeutic target for renal cell carcinoma.
European Review for Medical and Pharmacological Sciences **22**:46–54
DOI 10.26355/eurrev_201801_14099.

Huang Y, Liu D, Chen B, Zeng J, Wang L, Zhang S, Mo X, Li W. 2012. Loss of Bad expression
confers poor prognosis in non-small cell lung cancer. *Medical Oncology* **29(3)**:1648–1655
DOI 10.1007/s12032-011-0060-4.

Imrichová H, Hulselmans G, Atak ZK, Potier D, Aerts S. 2015. i-cisTarget 2015 update:
generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Research*
43(W1):W57–W64 DOI 10.1093/nar/gkv395.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics.
CA: A Cancer Journal for Clinicians **61(2)**:69–90 DOI 10.3322/caac.20107.

Jin S, Deng Y, Hao J-W, Li Y, Liu B, Yu Y, Shi FD, Zhou Q-H. 2014. NK cell phenotypic
modulation in lung cancer environment. *PLOS ONE* **9(10)**:e109976
DOI 10.1371/journal.pone.0109976.

Kikutake C, Yahara K. 2016. Identification of epigenetic biomarkers of lung adenocarcinoma
through multi-omics data analysis. *PLOS ONE* **11(4)**:e0152918
DOI 10.1371/journal.pone.0152918.

**Kim S, Herazo-Maya JD, Kang DD, Juan-Guardela BM, Tedrow J, Martinez FJ, Sciurba FC,
Tseng GC, Kaminski N. 2015.** Integrative phenotyping framework (iPF): integrative
clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics*
16(1):924 DOI 10.1186/s12864-015-2170-4.

Kooperstein SI, Schifrin A, Leahy TJ. 1965. Lung cancer and environment: a study of long-term
exposure to motor fumes in tunnels. *Journal of Occupational and Environmental Medicine*
7(12):626–629 DOI 10.1097/00043764-196512000-00005.

Li B, Cui Y, Diehn M, Li R. 2017. Development and validation of an individualized immune
prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncology*
3(11):1529–1537 DOI 10.1001/jamaoncol.2017.1609.

Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. 2014. starBase v2.0: decoding miRNA–ceRNA,
miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data.
Nucleic Acids Research **42(D1)**:D92–D97 DOI 10.1093/nar/gkt1248.

Li Y, Zhang W, Liu P, Xu Y, Tang L, Chen W, Guan X. 2018. Long non-coding RNA FENDRR
inhibits cell proliferation and is associated with good prognosis in breast cancer. *OncoTargets
and Therapy* **11**:1403–1412 DOI 10.2147/OTT.S149511.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion
for RNA-seq data with DESeq2. *Genome Biology* **15(12)**:550 DOI 10.1186/s13059-014-0550-8.

Lu Q-C, Rui Z-H, Guo Z-L, Xie W, Shan S, Ren T. 2018. LncRNA-DANCR contributes
to lung adenocarcinoma progression by sponging miR-496 to modulate mTOR expression.
Journal of Cellular and Molecular Medicine **22(3)**:1527–1537 DOI 10.1111/jcmm.13420.

Lv J, Xu L, Xu Y, Qiu M, Yang X, Wang J, Yin R, Xu L. 2014. Expression of MiRNA-221 in
non-small cell lung cancer tissues and correlation with prognosis. *Zhongguo Fei Ai Za Zhi*
17:221–225 DOI 10.3779/j.issn.1009-3419.2014.03.07.

Mattern J, Koomägi R, Volm M. 2002. Characteristics of long-term survivors of untreated
Lung Cancer. Lung Cancer **36(3)**:277–282 DOI 10.1016/S0169-5002(02)00010-7.

Men C-D, Liu Q-N, Ren Q. 2018. A prognostic 11 genes expression model for ovarian cancer. *Journal of Cellular Biochemistry* **119(2)**:1971–1978 DOI 10.1002/jcb.26358.

Miura H, Kato H, Konaka C, Usuda J, Uchida O, Taira O. 1996. Primary lung cancer of the middle lobe. Is its prognosis poor? *Lung Cancer* **14(2–3)**:273–279 DOI 10.1016/0169-5002(96)00553-3.

Onganer PU, Seckl MJ, Djamgoz MBA. 2005. Neuronal characteristics of small-cell lung cancer. *British Journal of Cancer* **93(11)**:1197–1201 DOI 10.1038/sj.bjc.6602857.

Phelps M, Coss C, Wang H, Cook M, Reproducibility Project: Cancer Biology. 2016. Registered report: coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *eLife* **5**:7154 DOI 10.7554/eLife.12470.

García Prim JM, González Barcala FJ, Paz Esquete J, Pose Reino A, Fondevila López A, Valdés Cuadrado L. 2010. Lung cancer in a health area of Spain: incidence, characteristics and survival. *European Journal of Cancer Care* **19(2)**:227–233 DOI 10.1111/j.1365-2354.2008.01008.x.

Pujol J-L, Monnier A, Berille J, Cerrina M-L, Douillard J-Y, Rivière A, Grandgirard A, Gouva S, Bizzari J-P, Le Chevalier T. 1994. Phase II study of nitrosourea fotemustine as single-drug chemotherapy in poor-prognosis non-small-cell lung cancer. *British Journal of Cancer* **69(6)**:1136–1140 DOI 10.1038/bjc.1994.223.

Putra AC, Eguchi H, Lee KL, Yamane Y, Gustine E, Isobe T, Nishiyama M, Hiyama K, Poellinger L, Tanimoto K. 2015. The A Allele at rs13419896 of EPAS1 is associated with enhanced expression and poor prognosis for non-small cell lung cancer. *PLOS ONE* **10(8)**:e0134496 DOI 10.1371/journal.pone.0134496.

Reinmuth N, Stumpf P, Stumpf A, Muley T, Kobinger S, Hoffmann H, Herth FJF, Schnabel PA, Bischoff H, Thomas M. 2014. Characteristics of lung cancer after a previous malignancy. *Respiratory Medicine* **108(6)**:910–917 DOI 10.1016/j.rmed.2014.02.015.

R Core Team. 2017. *R: A language and environment for statistical computing.* Version 3.4.3. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/.*

Ringbaek T, Borgeskov S, Lange P, Viskum K. 1999. Diagnostic and therapeutic process and prognosis in suspected lung cancer. *Scandinavian Cardiovascular Journal* **33(6)**:337–343 DOI 10.1080/14017439950141399.

Shang J, Song Q, Yang Z, Li D, Chen W, Luo L, Wang Y, Yang J, Li S. 2017. Identification of lung adenocarcinoma specific dysregulated genes with diagnostic and prognostic value across 27 TCGA cancer types. *Oncotarget* **8(50)**:87292–87306 DOI 10.18632/oncotarget.19823.

Shukla GC, Singh J, Barik S. 2011. MicroRNAs: processing, maturation, target recognition and regulatory functions. *Molecular and Cellular Pharmacology* **3**:83–92.

Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K, Zhu J. 2018. Five key lncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. *Journal of Cellular Biochemistry* **119(6)**:4559–4569 DOI 10.1002/jcb.26598.

Tamura T, Kurishima K, Watanabe H, Shiozawa T, Nakazawa K, Ishikawa H, Satoh H, Hizawa N. 2015. Characteristics of clinical no metastatic non-small cell lung cancer. *Lung Cancer* **89(1)**:71–75 DOI 10.1016/j.lungcan.2015.04.002.

Tamura M, Sasaki Y, Koyama R, Takeda K, Idogawa M, Tokino T. 2014. Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. *Oncogene* **33(40)**:4837–4846 DOI 10.1038/onc.2013.427.

Tang R-X, Chen W-J, He R-Q, Zeng J-H, Liang L, Li S-K, Ma J, Luo D-Z, Chen G. 2017. Identification of a RNA-Seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. *Oncotarget* **8(31)**:50761–50773 DOI 10.18632/oncotarget.17098.

Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, Lieberman J, Rigoutsos I, Pandolfi PP. 2011. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147(2)**:344–357 DOI 10.1016/j.cell.2011.09.029.

Tay Y, Rinn J, Pandolfi PP. 2014. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505(7483)**:344–352 DOI 10.1038/nature12986.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. 2015. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* **65(2)**:87–108 DOI 10.3322/caac.21262.

Truong MT, Viswanathan C, Erasmus JJ. 2011. Positron emission tomography/computed tomography in lung cancer staging, prognosis, and assessment of therapeutic response. *Journal of Thoracic Imaging* **26(2)**:132–146 DOI 10.1097/RTI.0b013e3182128704.

Wang W, Wu F, Zhao Z, Wang KY, Huang RY, Wang HY, Lan Q, Wang JF, Zhao JZ. 2018. Long noncoding RNA LINC00152 is a potential prognostic biomarker in patients with high-grade glioma. *CNS Neuroscience & Therapeutics* **24(10)**:957–966 DOI 10.1111/cns.12850.

Weiss W. 1983. Environment and type of lung cancer. *American Review of Respiratory Disease* **128**:1099–1100 DOI 10.1164/arrd.1983.128.6.1099a.

Wuttig D, Zastrow S, Füssel S, Toma MI, Meinhardt M, Kalman K, Junker K, Sanjmyatav J, Boll K, Hackermüller J, Rolle A, Grimm M-O, Wirth MP. 2012. CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. *International Journal of Cancer* **131(5)**: E693–E704 DOI 10.1002/ijc.27419.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* **39(suppl_2)**:W316–W322 DOI 10.1093/nar/gkr483.

Xu T-P, Huang M-D, Xia R, Liu X-X, Sun M, Yin L, Chen W-M, Han L, Zhang E-B, Kong R, De W, Shu Y-Q. 2014. Decreased expression of the long non-coding RNA FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *Journal of Hematology & Oncology* **7(1)**:63 DOI 10.1186/s13045-014-0063-7.

Yan H, Bi L, Wang Y, Zhang X, Hou Z, Wang Q, Snijders AM, Mao J-H. 2017. Integrative analysis of multi-omics data reveals distinct impacts of DDB1-CUL4 associated factors in human lung adenocarcinomas. *Scientific Reports* **7(1)**:333 DOI 10.1038/s41598-017-00512-1.

Yin Z, Xu M, Li P. 2017. Acts as an oncogenic role by directly targeting TIMP2 in non-small-cell lung carcinoma. *Gene* **620**:46–53 DOI 10.1016/j.gene.2017.04.007.

Yu L, Fang F, Lu S, Li X, Yang Y, Wang Z. 2017. lncRNA-HIT promotes cell proliferation of non-small cell lung cancer by association with E2F1. *Cancer Gene Therapy* **24(5)**:221–226 DOI 10.1038/cgt.2017.10.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4(1)**:Article17 DOI 10.2202/1544-6115.1128.

Zhen Q, Gao L-N, Wang R-F, Chu W-W, Zhang Y-X, Zhao X-J, Lv B-L, Liu J-B. 2018. LncRNA DANCR promotes lung cancer by sequestering miR-216a. *Cancer Control* **25(1)**:1073274818769849 DOI 10.1177/1073274818769849.