

# Bioinformatics analysis of microarray data to identify the candidate biomarkers of lung adenocarcinoma

Tingting Guo\*, Hongtao Ma\* and Yubai Zhou

Department of Biotechnology, College of Life Science & Bioengineering, Beijing University of Technology, Beijing, China

\* These authors contributed equally to this work.

## ABSTRACT

**Background:** Lung adenocarcinoma (LUAD) is the major subtype of lung cancer and the most lethal malignant disease worldwide. However, the molecular mechanisms underlying LUAD are not fully understood.

**Methods:** Four datasets (GSE118370, GSE85841, GSE43458 and GSE32863) were obtained from the gene expression omnibus (GEO). Identification of differentially expressed genes (DEGs) and functional enrichment analysis were performed using the limma and clusterProfiler packages, respectively. A protein–protein interaction (PPI) network was constructed via Search Tool for the Retrieval of Interacting Genes (STRING) database, and the module analysis was performed by Cytoscape. Then, overall survival analysis was performed using the Kaplan–Meier curve, and prognostic candidate biomarkers were further analyzed using the Oncomine database.

**Results:** Totally, 349 DEGs were identified, including 275 downregulated and 74 upregulated genes which were significantly enriched in the biological process of extracellular structure organization, leukocyte migration and response to peptide. The mainly enriched pathways were complement and coagulation cascades, malaria and prion diseases. By extracting key modules from the PPI network, 11 hub genes were screened out. Survival analysis showed that except VSIG4, other hub genes may be involved in the development of LUAD, in which MYH10, METTL7A, FCER1G and TMOD1 have not been reported previously to correlated with LUAD. Briefly, novel hub genes identified in this study will help to deepen our understanding of the molecular mechanisms of LUAD carcinogenesis and progression, and to discover candidate targets for early detection and treatment of LUAD.

**Subjects** Bioinformatics, Oncology, Data Mining and Machine Learning

**Keywords** Lung adenocarcinoma, Differentially expressed genes, Microarray, Biomarker, Bioinformatical analysis

## INTRODUCTION

Lung cancer remains the one of leading healthy issues worldwide, with as estimated 2.1 million new cases and 1.8 million deaths in 2018 (Bray *et al.*, 2018). It has been ranked the first and second cancer morbidity of male and female in China, respectively, and has the highest mortality rate (Sun *et al.*, 2018). Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer (Maemura *et al.*, 2018; Walters *et al.*, 2013). More than 60%

Submitted 9 April 2019

Accepted 18 June 2019

Published 10 July 2019

Corresponding author

Yubai Zhou, zhoyubai@bjut.edu.cn

Academic editor

Yonghong Peng

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.7313

© Copyright  
2019 Guo *et al.*

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

of LUAD patients were observed harboring targetable gene alterations, which leading to remarkable responses in treating with tyrosine kinase inhibitors, and associated with improved survival rate (*Kris et al., 2014*). Despite the substantial advance in combined therapies, the prognosis of LUAD is still dismal, the 5-year survival rate is not over 20% (*Chen et al., 2014b; Ettinger et al., 2013*). Lacking sensitive and specific early biomarkers, a high possibility of drug resistance and metastasis is considered to contribute the high mortality of this disease. Therefore, there has a pressing need for identifying the more sensitive and specific biomarkers or drug targets of LUAD for developing effective diagnosis and treatment strategies.

Microarray technology provides an all-in-one system biology solution from hardware to software systems. It can simultaneously scan the hybridization signals of tens of thousands of gene probes in the chip and carry out quantitative analysis on the transcriptome profile of samples. Recent advances especially in the algorithms of probe signal detection and analysis, such as the introduction of artificial intelligence technologies, will make the results of microarray more accurate and reliable (*Gan et al., 2019a, 2019b; Peng, 2006*). The microarray technique also provides a powerful tool for exploring the gene regulation pattern and molecular mechanisms involved in oncogenesis and progression of LUAD. Recently, different types of biomarkers including coding genes, miRNAs, long non-coding RNAs and circRNAs have been identified in lung cancer. Dysregulation of these molecules is involved in the tumor progression or is associated with the prognosis of patients (*Di et al., 2019; Vargas & Harris, 2016; Vencken, Greene & McKiernan, 2015; Wei & Zhou, 2016*). In view of the complexity of the molecular regulatory network of LUAD, current studies on tumor biomarkers are not sufficient. Therefore, it is still necessary to identify novel prognostic biomarkers, which will help us develop more sensitive and effective diagnostic and therapeutic strategies. However, limited sample size and significant variability among different projects make it hard to obtain credible results. In this study, four microarray datasets containing mRNA expression data between LUAD and non-cancerous tissues were downloaded from Gene Expression Omnibus (GEO) and the differentially expressed genes (DEGs) were screened out. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein-protein interaction (PPI) network analyses were performed to explore the key modules and hub genes involved in LUAD progression. In sum, 349 DEGs and 10 hub genes were screened out, which may be candidate biomarkers for LUAD.

## MATERIALS AND METHODS

### Data download and pre-processing

Four datasets (*GSE118370, GSE32863, GSE85841 and GSE43458*) which contain the gene expression data of LUAD and normal tissues, were downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) by getGEO function in package GEOquery (*Davis & Meltzer, 2007*). The detail information of GEO datasets was listed in [Table 1](#). The raw expression files of four microarray datasets were pre-processed according to the method described previously with minor modifications (*Giulietti et al., 2016*). Briefly, the CEL format files were input and background correction and normalization were

**Table 1** The detail information of four GEO datasets.

ID	Tissue	Platform	Normal	Tumor
GSE118370	LUAD	GPL570	6	6
GSE85841	LUAD	GPL20115	8	8
GSE43458	LUAD	GPL6244	30 (never-smoker)	40 (never-smoker)
GSE32863	LUAD	GPL6884	58	58

**Note:**

GEO, Gene Expression Omnibus; LUAD, lung adenocarcinoma.

conducted using the Robust multichip average function implemented in affy package in R environment (Bolstad *et al.*, 2003; Irizarry *et al.*, 2003). Next, the array probes were converted into matched gene symbols according to annotation information. In case of multiple probes corresponding to a single gene, the value of gene expression was designated as the mean of the probes. Then, the batch effects among different platforms were removed by ComBat function of the sva package (Leek *et al.*, 2012). Finally, the normalized microarray-based data of four datasets were merged into a single global dataset which contained a total of 12,926 common genes in all four GEO datasets.

### Identification of DEGs

Identifying DEGs in different disease states and investigating their functions and interactions may help to unravel potential regulatory mechanisms for disease occurrence and progression. In the present study, the DEGs between LUAD and normal tissues were identified by limma package (Ritchie *et al.*, 2015). The Benjamini–Hochberg procedure was introduced to reduce the false positive rate (FDR) in multiple comparisons (Benjamini & Hochberg, 1995). Genes with  $|\log_2 \text{Fold Change}| \geq 1$  and  $\text{FDR} < 0.05$  were considered as DEGs.

### GO and KEGG enrichment analysis

Gene Ontology and KEGG analyses were conducted using enrichGO and enrichKEGG functions of clusterProfiler package, respectively (Yu *et al.*, 2012).  $p.\text{adjust} (\text{FDR}) < 0.05$  was considered to be statistically significant.

### Construction of PPI network and module analysis

The network of proteins interaction provides valuable clues for understanding the molecular mechanisms underlying the progress of carcinoma. The PPI network was constructed by Search Tool for the Retrieval of Interacting Genes (STRING) (<https://string-db.org/>) with interaction score of 0.9 as the threshold. Subsequently, the candidate modules were detected by Cytoscape plugin molecular complex detection (MCODE) with default parameters: degree cut-off = 2, node score cut-off = 0.2,  $k\text{-core} = 2$ , and max depth = 100.

### Hub gene analysis

The seed genes in modules referred to hub genes. The overall survival analyses were performed using online tool Kaplan–Meier Plotter (<http://kmplot.com/>) (Gyorffy *et al.*, 2013). The logrank  $P < 0.05$  was considered statistically significant. The association of

**Table 2** Top ten up- and down-regulated DEGs.

Up-regulated DEGs			Down-regulated DEGs		
Gene symbol	Log2FC	FDR	Gene symbol	Log2FC	FDR
SPP1	3.08474	1.74E-33	FABP4	-3.37555	1.25E-47
OCIAD2	1.477746	1.13E-29	STX11	-2.01766	1.51E-44
ETV4	1.366623	5.38E-27	CAV1	-2.70539	6.04E-43
TOP2A	1.777989	3.24E-26	FHL1	-2.3099	5.77E-42
COL10A1	1.364488	3.74E-26	TEK	-2.30698	1.09E-40
PROM2	1.570558	7.77E-26	AGER	-2.69928	6.05E-40
MMP11	1.929847	1.03E-23	FMO2	-2.44791	1.42E-39
UBE2T	1.154061	5.78E-23	CRYAB	-1.98382	1.77E-39
ABCC3	1.444218	3.71E-22	GRK5	-1.4854	1.53E-38
BAIAP2L1	1.067336	4.15E-22	TMEM100	-2.928	4.79E-38

**Note:**

DEGs, differentially expressed genes; FDR, false discovery rate; Log2FC, Log2 (Fold Change).

expression level of hub genes with clinical traits were analysis using the Oncomine database (*Rhodes et al., 2007*).

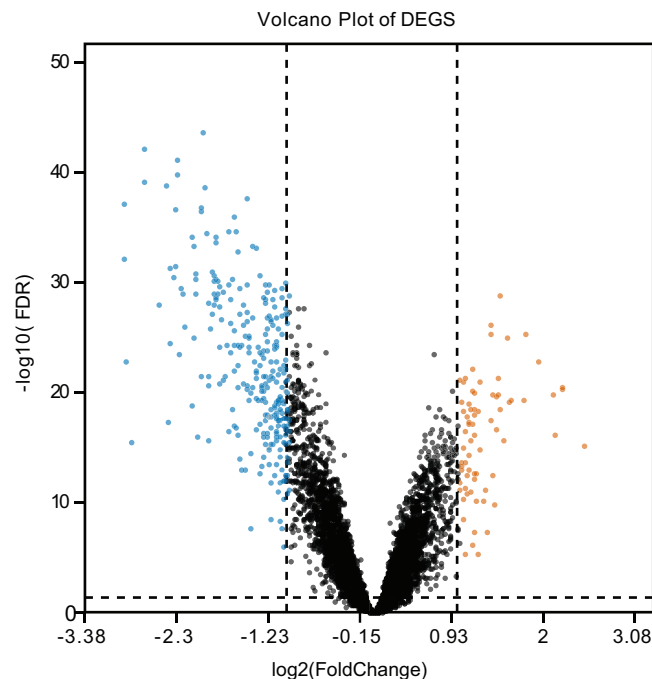
## RESULTS

### Data preprocessing and DEG screening

Four GEO datasets were downloaded, pre-processed and merged into a global dataset which contained 112 LUAD and 102 normal samples. Totally, 349 DEGs were identified by limma package (*Ritchie et al., 2015*), including 74 up-regulated genes and 275 down-regulated genes. The most statistically significant up-regulated and down-regulated genes are listed in [Table 2](#). The distribution of DEGs was presented by volcano plot ([Fig. 1](#)).

### GO and KEGG analysis

The biological functions and pathways analyses were conducted using R package clusterProfiler (*Yu et al., 2012*). The GO categories of biological process (BP), cellular component (CC) and molecular function (MF) were enriched respectively ([Figs. 2A–2C](#)) and the top 15 GO terms of the up-regulated and down-regulated DEGs were listed in [Tables S1](#) and [S2](#), respectively. The up-regulated DEGs were mainly associated with extracellular matrix (ECM) processing, such as ECM organization (BP, [GO:0030198](#)), ECM (CC, [GO:0031012](#)) and serine-type endopeptidase activity (MF, [GO:0004252](#)). The down-regulated DEGs were most significantly related to response to corticosteroid (BP, [GO:0031960](#)), ECM (CC, [GO:0031012](#)) and growth factor binding (MF, [GO:0019838](#)). The DEGs were mainly enriched in pathways of complement and coagulation cascades ([hsa04610](#)), malaria ([hsa05144](#)), prion diseases ([hsa05020](#)), fluid shear stress and atherosclerosis ([hsa05418](#)), AGE-RAGE signaling pathway in diabetic complications ([hsa04933](#)), vascular smooth muscle contraction ([hsa04270](#)), IL-17 signaling pathway ([hsa04657](#)), leukocyte transendothelial migration ([hsa04670](#)), protein digestion and absorption ([hsa04974](#)) and drug metabolism—cytochrome P450 ([hsa00982](#)) ([Table 3](#); [Fig. 2D](#)).



**Figure 1** Volcano plot of the DEGs. The vermilion and blue dots represent DEGs filtered based on the cut-off values of  $|\log_2\text{FoldChange}| > 1.0$  and adjusted  $P$ -value (FDR)  $< 0.05$ , while the black dots represent genes that are not satisfied the cut-off values of differential expression. The horizon dotted line indicates the position of  $-\log_{10}(\text{FDR}) = 0.05$ , and the vertical dotted lines indicate the positions of  $|\log_2\text{FoldChange}| = 1.0$ . Vermilion, upregulation; blue, downregulation. DEGs, differentially expressed genes; FDR, false discovery rate.

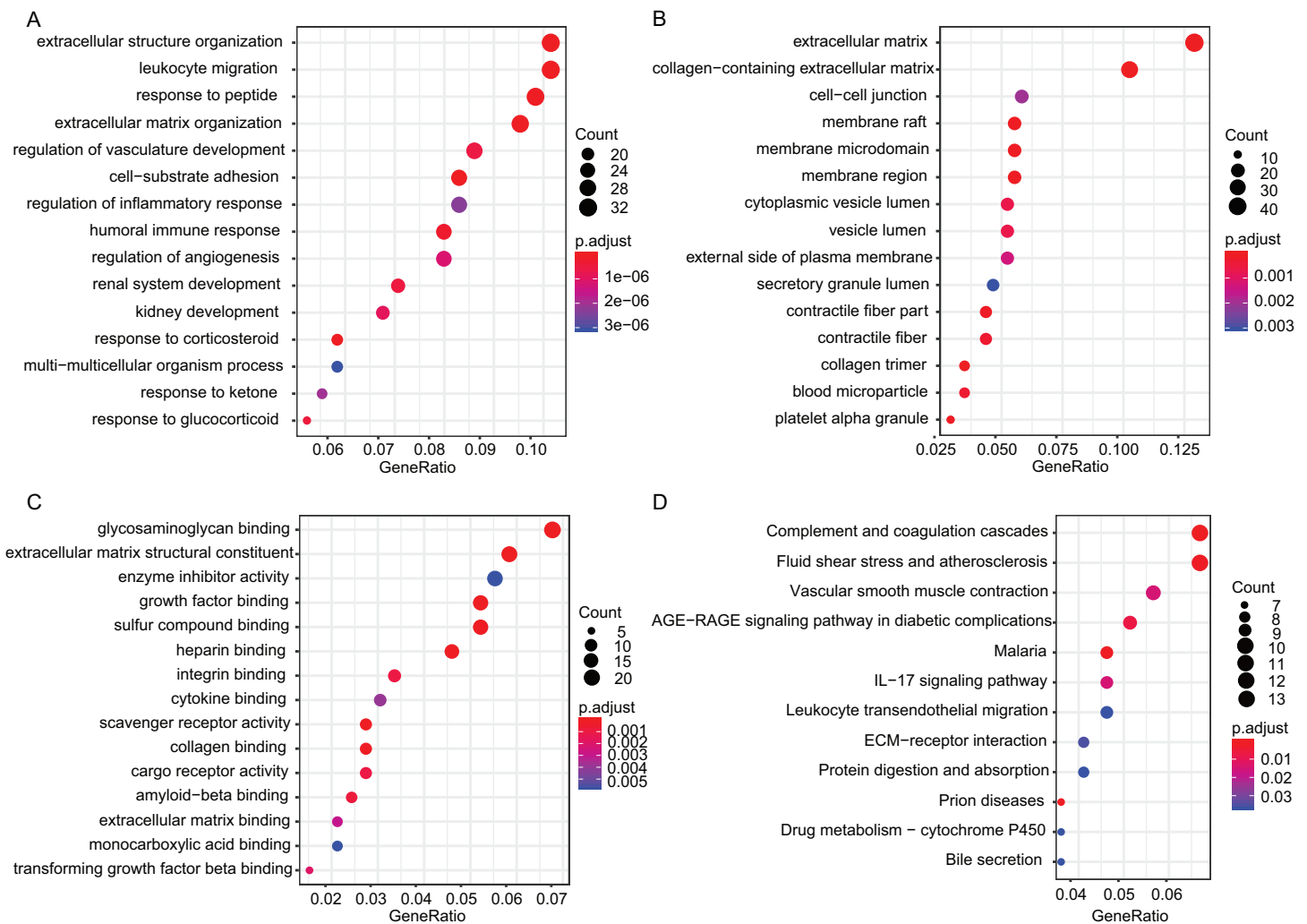
Full-size  DOI: [10.7717/peerj.7313/fig-1](https://doi.org/10.7717/peerj.7313/fig-1)

## Construction of PPI network and module analysis

Protein–protein interaction network reflect the spatiotemporal relationship of macromolecules within the cell which will provide valuable information about molecular mechanisms in physiological and pathological process. To explore the molecular mechanisms underlying LUAD progression, the online STRING database was applied to construct the PPI network. The interaction score of 0.9 (highest confidence) was set as threshold, and nodes without connections were removed from network. Finally, the PPI network consisted of 349 nodes with 277 edges, and average local clustering coefficient was 0.337 (PPI enrichment  $P$ -value  $< 1.0\text{E-}16$ ) (Fig. 3A). Then, the key modules were identified via MCODE plugin. A total of 11 functional clusters of modules and related hub genes were detected. The top three significant modules were presented in Figs. 3C–3E. The KEGG analysis of module genes revealed that the top three modules were mainly associated with the chemokine signaling pathway (hsa04062), complement and coagulation cascades (hsa04610), human cytomegalovirus infection (hsa05163) and vascular smooth muscle contraction (hsa04270) (Fig. 3B).

## Hub genes analysis

A total of 11 genes were identified as hub genes. The overall survival analysis of the hub genes was performed using Kaplan–Meier curve. Except VSIG4, LUAD patients with downregulated ADAMTS8, AOX1, EFEMP1, METTL7A, MYH10, PTGER4, TMOD1,



**Figure 2** GO and KEGG analysis of DEGs. (A–C) The top 15 terms of GO categories of biological process (BP), cellular component (CC) and molecular function MF, respectively. (D) KEGG pathway analysis of DEGs, p.adjust (FDR) < 0.05 was considered significantly. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate. [Full-size !\[\]\(1679558f37f6db0dd8360a2a7e913e90\_img.jpg\) DOI: 10.7717/peerj.7313/fig-2](https://doi.org/10.7717/peerj.7313/fig-2)

CDH13 and upregulated PRC1 showed worse overall survival (Figs. 4A–4I). It is worth noting that FCER1G is downregulated in LUAD patients, but the low expression level is associated with better overall survival (HR = 1.87) (Fig. 4J). Subsequently, the expression status of hub genes with HR < 0.5 or HR > 2 were further validated using the Oncomine database. The result showed that ADAMTS8, METTL7A and MYH10 were significantly downregulated and PRC1 was markedly overexpressed in LUAD in the different datasets (Fig. 5). In the Okayama Lung dataset, the alternation of ADAMTS8, METTL7A, MYH10 and PRC1 were associated with tumor grade (Fig. 6), implicating vital roles of these genes in the carcinogenesis or progression of LUAD.

## DISCUSSION

In this study, four GEO datasets were analyzed and 349 DEGs were identified, including 74 up-regulated and 275 down-regulated genes. The KEGG analysis revealed the top three



**Table 3** KEGG enriched pathways of DEGs.

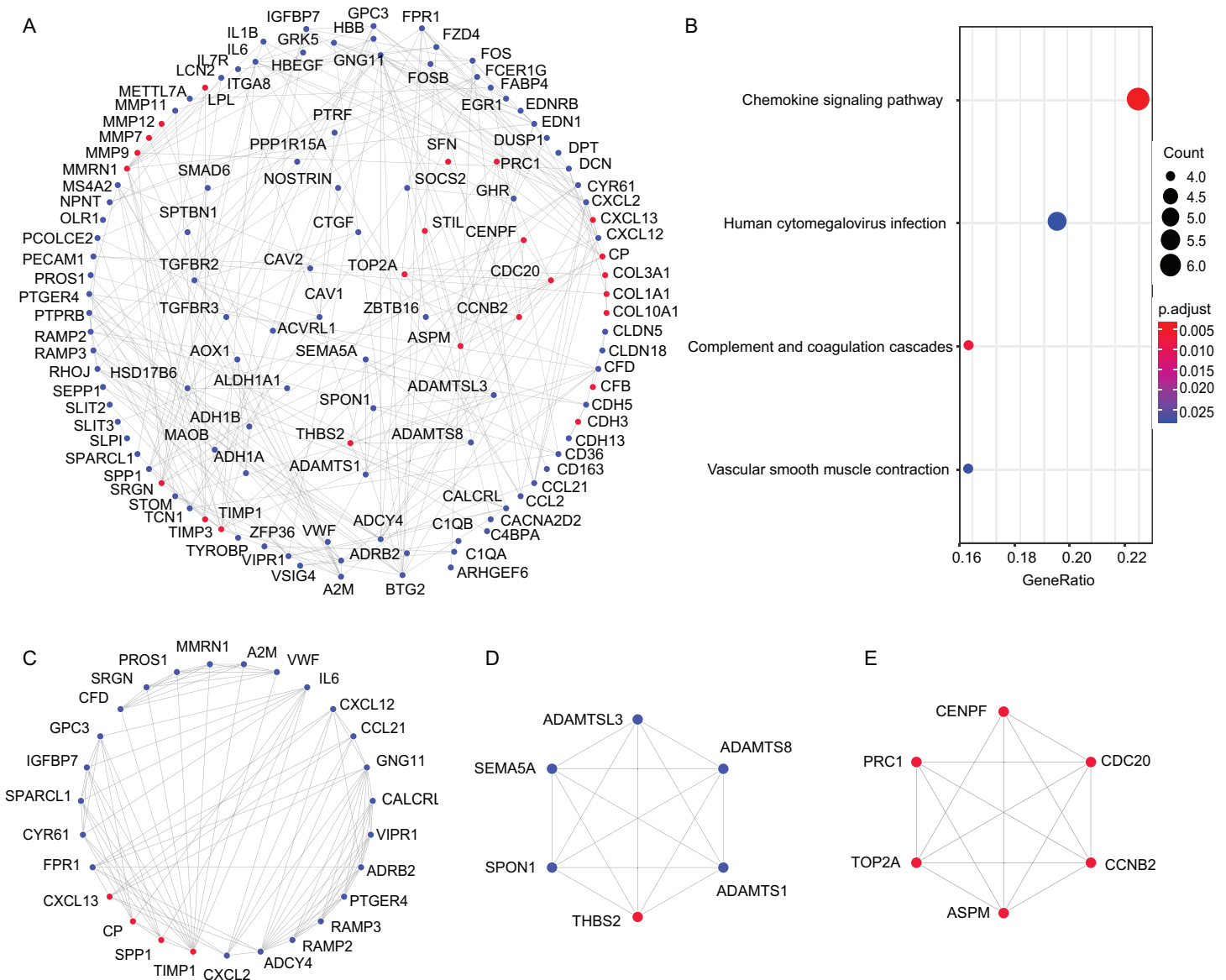
ID	Description	Count	p.adjust (FDR)
hsa04610	Complement and coagulation cascades	13	1.12E-05
hsa05144	Malaria	9	0.000271
hsa05020	Prion diseases	7	0.001406
hsa05418	Fluid shear stress and atherosclerosis	13	0.001921
hsa04933	AGE-RAGE signaling pathway in diabetic complications	10	0.007362
hsa04270	Vascular smooth muscle contraction	11	0.014433
hsa04657	IL-17 signaling pathway	9	0.014433
hsa04974	Protein digestion and absorption	8	0.039299
hsa04670	Leukocyte transendothelial migration	9	0.039299
hsa00982	Drug metabolism—cytochrome P450	7	0.039299

**Note:**

KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; FDR, false discovery rate.

enriched pathways were complement and coagulation cascades, malaria and prion diseases. These annotation results provided valuable clues to reveal molecular interactions in the development of LUAD. Indeed, the complement system has been reported to play critical roles in tumor progression (*Zhao et al., 2019*). The overexpressed C3 in blood and downregulated C3 in tumor tissues were observed in lung cancer patients, and related to poorer prognosis (*Ajona et al., 2013; Lin et al., 2014; Mehan et al., 2012*). The confusing experimental results suggest that the complement system may be involved in complex tumor regulatory processes by reshaping tumor microenvironment, which is worthy of further study. In recent years, malaria and prion disease, which had not previously attracted much attention, have also been found to be associated with tumors. Epidemiological study has shown that the incidence of malaria is negatively correlated with the mortality of colorectal cancer, breast cancer and lung cancer (*Qin et al., 2017*). The proposed anti-tumor mechanisms included systemly stimulating the immune responses (*Chen et al., 2011*) and inhibiting key pathways in tumor progress (*Deng et al., 2018*). Previously, prion protein (PrPc) was thought to play a role only in the central nervous system. However, accumulating evidence shows that PrPc has wider biological functions that were not previously expected (*Mehrpour & Codogno, 2010*). PrPc may be associated with the biology of many cancers, and overexpression of PrPc promotes the proliferation, invasion and metastasis of the gastric cancer cell (*Liang et al., 2009; Mehrpour & Codogno, 2010; Pan et al., 2006*). These data may provide new ideas and directions for the mechanism research and therapeutic strategy of LUAD. The GO analyses indicated that DEGs were significantly related to BP of ECM organization and process. Previous studies have reported that ECM remodeling promotes cancer progression and is associated with a poor prognosis in lung cancer patients (*Kopparam et al., 2017; Xia et al., 2012*). The up- and down-regulated DEGs were simultaneously enriched into ECM process, which is consistent with the propensity for metastasis and highly invasive characteristics of LUAD.

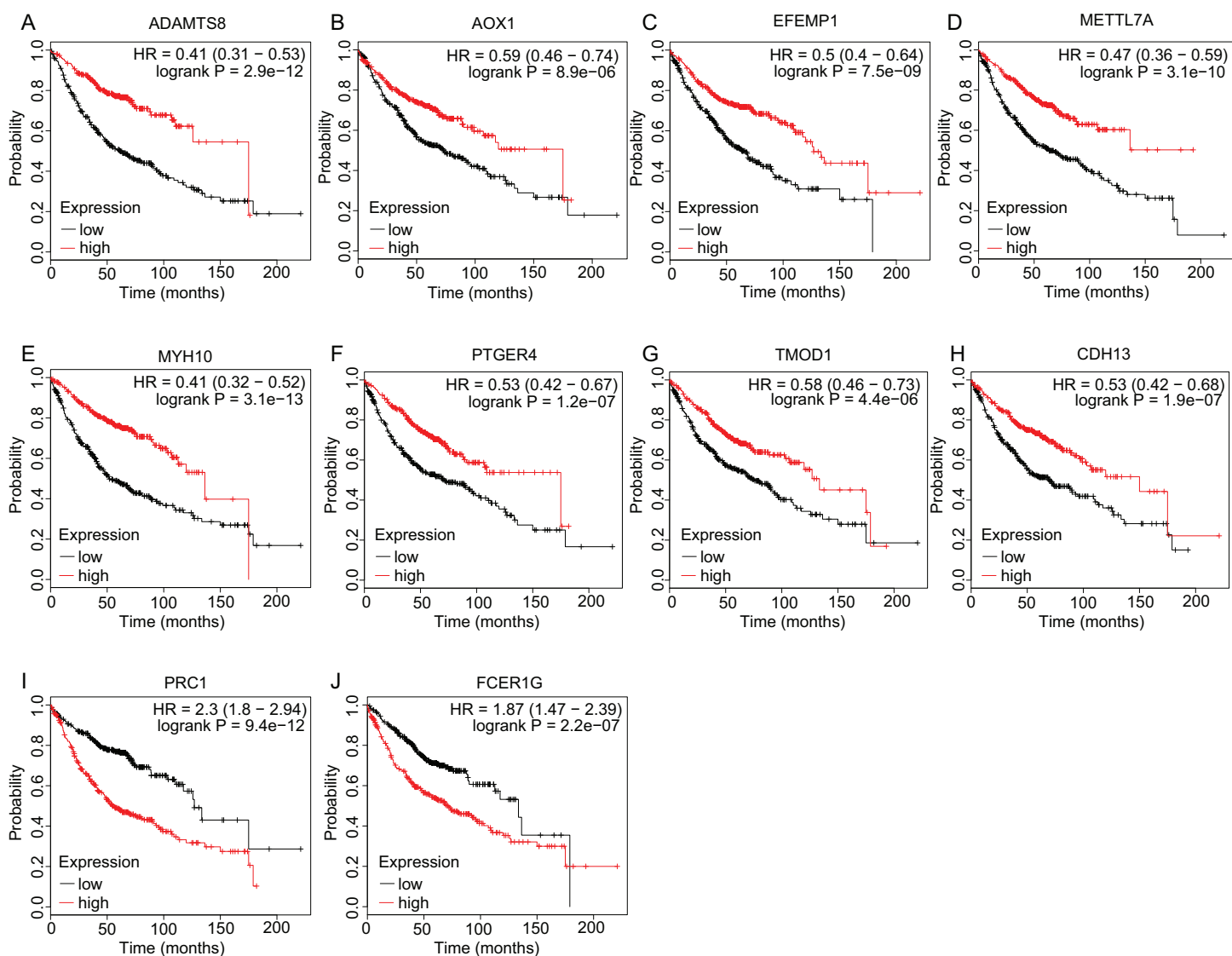
In line with the GO analysis, among 11 hub genes identified by PPI network and modules analysis, EFEMP1, PTGER4, ADAMTS8, CDH13, MYH10 and METTL7A are



**Figure 3 Protein-protein interaction network of DEGs and modules analysis.** (A) PPI network of DEGs; (B) The KEGG enrichment analysis of the genes in the top three modules. (C-E) the top 3 modules of PPI network. The red nodes represent the upregulated DEGs. The blue nodes represent the downregulated DEGs. PPI, protein-protein interaction; DEG, differentially expressed gene; KEGG, Kyoto Encyclopedia of Genes and Genomes. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02\_img.jpg\) DOI: 10.7717/peerj.7313/fig-3](https://doi.org/10.7717/peerj.7313/fig-3)

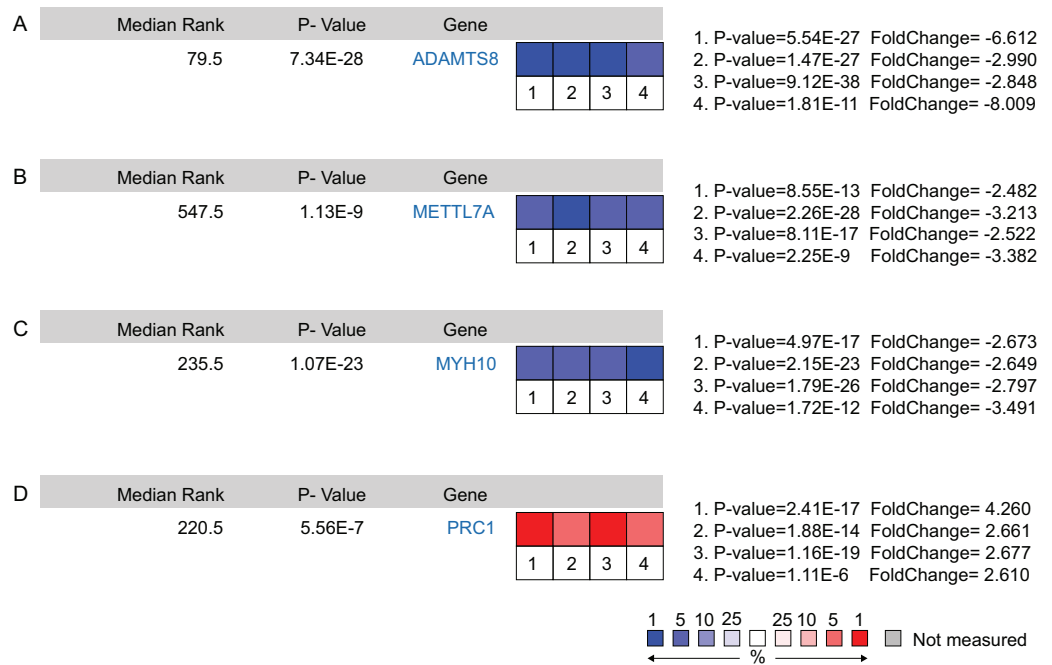
associated with ECM process, and survival analysis indicated that down-regulation of these hub genes was associated with worse overall survival. EFEMP1 plays distinct biological functions in different tumors. In osteosarcoma and gliomas, EFEMP1 is overexpressed and promotes the invasion and metastasis of tumor cells in vitro and in vivo by activating the expression of MMP-2 and notch signaling, respectively (*Hu et al., 2009, 2012; Wang et al., 2015*), while in gastric cancer, endometrial carcinoma, hepatocellular carcinoma and lung cancer, EFEMP1 is downregulated, and is proposed as a prognosis biomarker (*Chen et al., 2014a; Kim et al., 2014; Nomoto et al., 2010; Yang et al., 2013;*



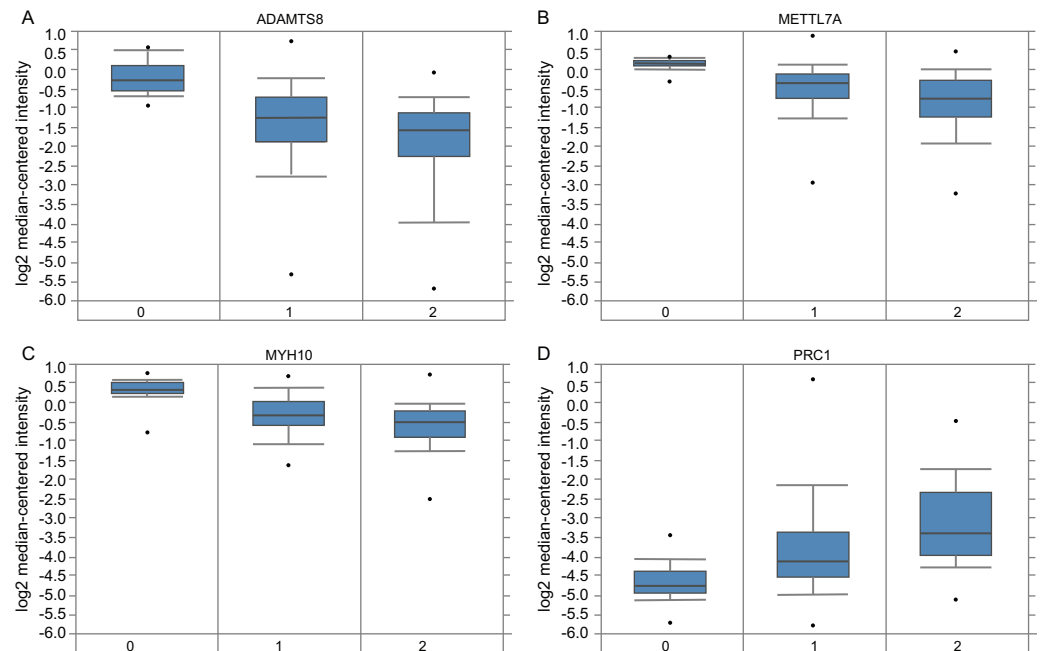


**Figure 4** Overall survival analyses of hub genes. (A–J) The overall survival analyses of hub genes were performed using Kaplan–Meier Plotter online platform. Logrank  $P < 0.05$  was considered statistically significant. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674\_img.jpg\) DOI: 10.7717/peerj.7313/fig-4](https://doi.org/10.7717/peerj.7313/fig-4)

*Yue et al., 2007; Zhu et al., 2014*). Overexpression of EFEMP1 has been reported to suppressed invasion and migration of LUAD cells via inhibiting the epithelial-to-mesenchymal transition pathway (*Kim et al., 2014*). PTGER4 is overexpressed and proposed as a therapeutic target for LUAD and other cancers (*Doherty et al., 2009; Fulton, Ma & Kundu, 2006; Heinrichs et al., 2018; Kim et al., 2010; Ma et al., 2006; Xin et al., 2012*). The above report is inconsistent with our results, which may be due to the heterogeneity of the tumor and the limited number of samples. Therefore, subsequent large sample functional verification is required. ADAMTS8 encodes an inactive proenzyme and forms mature active enzyme by proteolysis (*Apte, 2009*). ADAMTS8 is identified as a secretory angiogenesis inhibitor that inhibits VEGF-mediated angiogenesis by blocking the EGFR signaling pathway (*Choi et al., 2014; Dunn et al., 2006; Vazquez et al., 1999*).



**Figure 5** Oncomine analysis of cancer vs. normal tissue of ADAMTS8, METTL7A, MYH10 and PRC1. Heat maps of ADAMTS8 (A), METTL7A (B), MYH10 (C) and PRC1 (D) gene expression in lung adenocarcinoma samples vs. normal tissues. (1) Lung Adenocarcinoma vs. Normal, Hou Lung (Hou et al., 2010); (2) Lung Adenocarcinoma vs. Normal, Landi Lung (Landi et al., 2008); (3) Lung Adenocarcinoma vs. Normal, Selamat Lung (Selamat et al., 2012); (4) Lung Adenocarcinoma vs. Normal, Su Lung (Su et al., 2007). Full-size [DOI: 10.7717/peerj.7313/fig-5](https://doi.org/10.7717/peerj.7313/fig-5)



**Figure 6** The association between the expression of selected hub genes and tumor stage. (A–D) The expressions of ADAMTS8, METTL7A, MYH10, PRC1 were correlated with tumor stage in the Okayama Lung dataset. 0: No value; 1: Stage I; 2: Stage II. Full-size [DOI: 10.7717/peerj.7313/fig-6](https://doi.org/10.7717/peerj.7313/fig-6)

Downregulated ADAMTS8 in some cancers including NSCLC (*Heighway et al., 2002; Huang et al., 2019; Masui et al., 2001; Porter et al., 2004; Rodriguez-Rodero et al., 2013; Zhao et al., 2018*) is associated with poorer prognosis (*Drilon et al., 2014; Li et al., 2015; Porter et al., 2006*). CDH13 encodes a member of cadherin superfamily. The hypermethylation in promoter region of CDH13 was frequently observed in lung cancer, and proposed to correlate to drug sensitivity and poorer prognosis (*Kontic et al., 2012; Toyooka et al., 2006; Zhai & Li, 2014; Zhong et al., 2015*). As the only upregulated hub gene, PRC1 is involved in the process of cytokinesis and is upregulated in breast cancer, hepatocellular carcinoma, and lung cancer (*Jiang et al., 1998; Liu et al., 2018; Shimo et al., 2007; Zhan et al., 2017a*). Recent research has found that it promoted proliferation and metastasis of LUAD cells via potentiating the Wnt/ $\beta$ -catenin pathway (*Zhan et al., 2017b*), and inhibiting the expression of PRC1 in LUAD cells by miR-1-3p was reported to suppress tumorigenesis (*Li et al., 2019*).

Literature retrieval showed that the relation of LUAD and hub genes MYH10, METTL7A, FCER1G and TMOD1 has not been reported. MYH10 belongs to the myosin superfamily, which can regulate ECM remodeling (*Kim et al., 2018; Kim, Kurie & Ahn, 2015*). Methyltransferase METTL7A was found to facilitate Hepatitis C Virus propagation via recruitment of NS4B (*Park et al., 2015*). FCER1G encodes a high affinity IgE receptor which is associated with prognosis of renal clear cell carcinoma (*Chen et al., 2017*). As an aldehyde oxidase, TMOD1 is an actin-capping protein involved in regulation of the length and depolymerization of actin filaments. Overexpression of TMOD1 promotes cell proliferation of breast cancers and metastasis of oral squamous cell carcinoma (*Ito-Kureha et al., 2015; Suzuki et al., 2016*). However, the molecular mechanism of these newly identified hub genes in LUAD remains largely unknown, and further functional studies were warranted.

In present study, we have identified a set of candidate biomarkers that may play an important role in the progression of LUAD. These newly identified hub genes could be used as research subjects for exploring their roles in the disease process, so as to further deepen our understanding of the molecular mechanism of LUAD, and also as potential prognostic biomarkers for clinical validation studies to clarify their prognostic effects. However, there are still some limitations in this study. First, this study is based on bioinformatics analysis of published data and lacks experimental verification, and the study cannot determine whether there is a causal relationship between the differential expression of hub genes and disease progression. Finally, although we combined four GEO datasets, the number of samples is still relatively small, which may lead to potential unreliable results. Therefore, subsequent bioinformatics analysis and experimental verification with larger samples are necessary.

## CONCLUSIONS

In summary, this study identified several DEGs by integrating four GEO datasets and extracted 11 hub genes from PPI network, among which 10 hub genes were shown to be related to the occurrence and development of LUAD and four hub genes have not been previously reported but may play an important role in LUAD. The molecular mechanism

of these novel hub genes in LUAD is worthy of further study, and relevant prognostic model can also be constructed based on these genes for risk assessment, classification and prognostic judgment of patients with LUAD.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the Major National Research and Development Projects (2012ZX10001005-007), the National High Technology Research Program of China (2012AA02A404), and the State Key Laboratory for Infectious Disease Prevention and Control (2011SKLID103). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Major National Research and Development Projects: 2012ZX10001005-007.

National High Technology Research Program of China: 2012AA02A404.

State Key Laboratory for Infectious Disease Prevention and Control: 2011SKLID103.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Tingting Guo performed the experiments, analyzed the data, prepared figures and/or tables.
- Hongtao Ma performed the experiments, analyzed the data, prepared figures and/or tables.
- Yubai Zhou 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.

### Data Availability

The following information was supplied regarding data availability:

Data is available at NCBI GEO under accession numbers [GSE118370](#), [GSE32863](#), [GSE85841](#) and [GSE43458](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7313#supplemental-information>.

## REFERENCES

- Ajona D, Pajares MJ, Corrales L, Perez-Gracia JL, Agorreta J, Lozano MD, Torre W, Massion PP, De-Torres JP, Jantus-Lewintre E, Camps C, Zulueta JJ, Montuenga LM, Pio R. 2013. Investigation of complement activation product c4d as a diagnostic and prognostic

- biomarker for lung cancer. *JNCI: Journal of the National Cancer Institute* **105(18)**:1385–1393 DOI [10.1093/jnci/djt205](https://doi.org/10.1093/jnci/djt205).
- Apte SS. 2009.** A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *Journal of Biological Chemistry* **284(46)**:31493–31497 DOI [10.1074/jbc.R109.052340](https://doi.org/10.1074/jbc.R109.052340).
- Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* **57**:289–300.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003.** A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19(2)**:185–193 DOI [10.1093/bioinformatics/19.2.185](https://doi.org/10.1093/bioinformatics/19.2.185).
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018.** Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68(6)**:394–424 DOI [10.3322/caac.21492](https://doi.org/10.3322/caac.21492).
- Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. 2014b.** Non-small-cell lung cancers: a heterogeneous set of diseases. *Nature Reviews Cancer* **14(8)**:535–546 DOI [10.1038/nrc3775](https://doi.org/10.1038/nrc3775).
- Chen L, He Z, Qin L, Li Q, Shi X, Zhao S, Chen L, Zhong N, Chen X. 2011.** Antitumor effect of malaria parasite infection in a murine Lewis lung cancer model through induction of innate and adaptive immunity. *PLOS ONE* **6(9)**:e24407 DOI [10.1371/journal.pone.0024407](https://doi.org/10.1371/journal.pone.0024407).
- Chen X, Meng J, Yue W, Yu J, Yang J, Yao Z, Zhang L. 2014a.** Fibulin-3 suppresses Wnt/ $\beta$ -catenin signaling and lung cancer invasion. *Carcinogenesis* **35(8)**:1707–1716 DOI [10.1093/carcin/bgu023](https://doi.org/10.1093/carcin/bgu023).
- Chen L, Yuan L, Wang Y, Wang G, Zhu Y, Cao R, Qian G, Xie C, Liu X, Xiao Y, Wang X. 2017.** Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. *International Journal of Biological Sciences* **13(11)**:1361–1372 DOI [10.7150/ijbs.21657](https://doi.org/10.7150/ijbs.21657).
- Choi GCG, Li J, Wang Y, Li L, Zhong L, Ma B, Su X, Ying J, Xiang T, Rha SY, Yu J, Sung JJ, Tsao SW, Chan AT, Tao Q. 2014.** The metalloprotease ADAMTS8 displays antitumor properties through antagonizing EGFR–MEK–ERK signaling and is silenced in carcinomas by CpG methylation. *Molecular Cancer Research* **12(2)**:228–238 DOI [10.1158/1541-7786.MCR-13-0195](https://doi.org/10.1158/1541-7786.MCR-13-0195).
- Davis S, Meltzer PS. 2007.** GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* **23(14)**:1846–1847 DOI [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254).
- Deng XF, Zhou D, Liu QX, Zheng H, Ding Y, Xu WY, Min JX, Dai JG. 2018.** Plasmodium circumsporozoite protein suppresses the growth of A549 cells via inhibiting nuclear transcription factor  $\kappa$ B. *Oncology Letters* **15**:6585–6591 DOI [10.3892/ol.2018.8115](https://doi.org/10.3892/ol.2018.8115).
- Di X, Jin X, Li R, Zhao M, Wang K. 2019.** CircRNAs and lung cancer: biomarkers and master regulators. *Life Sciences* **220**:177–185 DOI [10.1016/j.lfs.2019.01.055](https://doi.org/10.1016/j.lfs.2019.01.055).
- Doherty GA, Byrne SM, Molloy ES, Malhotra V, Austin SC, Kay EW, Murray FE, Fitzgerald DJ. 2009.** Proneoplastic effects of PGE2 mediated by EP4 receptor in colorectal cancer. *BMC Cancer* **9(1)**:207 DOI [10.1186/1471-2407-9-207](https://doi.org/10.1186/1471-2407-9-207).
- Drilon A, Sugita H, Sima CS, Zauderer M, Rudin CM, Kris MG, Rusch VW, Azzoli CG. 2014.** A prospective study of tumor suppressor gene methylation as a prognostic biomarker in surgically resected stage I to IIIA non-small-cell lung cancers. *Journal of Thoracic Oncology* **9(9)**:1272–1277 DOI [10.1097/JTO.0000000000000256](https://doi.org/10.1097/JTO.0000000000000256).

- Dunn JR, Reed JE, du Plessis DG, Shaw EJ, Reeves P, Gee AL, Warnke P, Walker C. 2006. Expression of ADAMTS-8, a secreted protease with antiangiogenic properties, is downregulated in brain tumours. *British Journal of Cancer* 94(8):1186–1193 DOI 10.1038/sj.bjc.6603006.
- Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, D'Amico TA, Demmy TL, Govindan R, Grannis FW Jr, Grant SC, Horn L, Jahan TM, Komaki R, Kong FM, Kris MG, Krug LM, Lackner RP, Lennes IT, Loo BW Jr, Martins R, Otterson GA, Patel JD, Pinder-Schenck MC, Pisters KM, Reckamp K, Riely GJ, Rohren E, Shapiro TA, Swanson SJ, Tauer K, Wood DE, Yang SC, Gregory K, Hughes M. 2013. Non-small cell lung cancer, version 2.2013. *Journal of the National Comprehensive Cancer Network* 11(6):645–653 DOI 10.6004/jnccn.2013.0084.
- Fulton AM, Ma X, Kundu N. 2006. Targeting prostaglandin E EP receptors to inhibit metastasis: figure 1. *Cancer Research* 66(20):9794–9797 DOI 10.1158/0008-5472.CAN-06-2067.
- Gan Z, Zeng N, Zou F, Chen J, Du M, Liao L-C, Li H, Zhang Y. 2019a. Multilevel segmentation optimized by physical information for gridding of microarray images. *IEEE Access* 7:32146–32153 DOI 10.1109/ACCESS.2019.2900249.
- Gan Z, Zou F, Zeng N, Xiong B, Liao L, Li H, Luo X, Du M. 2019b. Wavelet denoising algorithm based on NDOA compressed sensing for fluorescence image of microarray. *IEEE Access* 7:13338–13346 DOI 10.1109/ACCESS.2019.2891759.
- Giulietti M, Occhipinti G, Principato G, Piva F. 2016. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. *Cellular Oncology* 39(4):379–388 DOI 10.1007/s13402-016-0283-7.
- Gyorffy B, Surowiak P, Budczies J, Lanczky A. 2013. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLOS ONE* 8(12):e82241 DOI 10.1371/journal.pone.0082241.
- Heighway J, Knapp T, Boyce L, Brennand S, Field JK, Betticher DC, Ratschiller D, Gugger M, Donovan M, Lasek A, Rickert P. 2002. Expression profiling of primary non-small cell lung cancer for target identification. *Oncogene* 21(50):7749–7763 DOI 10.1038/sj.onc.1205979.
- Heinrichs SKM, Hess T, Becker J, Hamann L, Vashist YK, Butterbach K, Schmidt T, Alakus H, Krasniuk I, Hoblinger A, Lingohr P, Ludwig M, Hagel AF, Schildberg CW, Veits L, Gyvyte U, Weise K, Schuller V, Bohmer AC, Schroder J, Gehlen J, Kreuser N, Hofer S, Lang H, Lordick F, Malfertheiner P, Moehler M, Pech O, Vassos N, Rodermann E, Izbicki JR, Kruschewski M, Ott K, Schumann RR, Vieth M, Mangold E, Gasenko E, Kupcinkas L, Brenner H, Grimminger P, Bujanda L, Sopena F, Espinel J, Thomson C, Perez-Aisa A, Campo R, Geijo F, Collette D, Bruns C, Messerle K, Gockel I, Nothen MM, Lippert H, Ridwelski K, Lanas A, Keller G, Knapp M, Leja M, Kupcinkas J, Garcia-Gonzalez MA, Venerito M, Schumacher J. 2018. Evidence for *PTGER4*, *PSCA*, and *MBOAT7* as risk genes for gastric cancer on the genome and transcriptome level. *Cancer Medicine* 7(10):5057–5065 DOI 10.1002/cam4.1719.
- Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA, Hoogsteden HC, Grosveld F, Philipsen S. 2010. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLOS ONE* 5(4):e10312 DOI 10.1371/journal.pone.0010312.
- Hu B, Nandhu MS, Sim H, Agudelo-Garcia PA, Saldivar JC, Dolan CE, Mora ME, Nuovo GJ, Cole SE, Viapiano MS. 2012. Fibulin-3 promotes glioma growth and resistance through a novel paracrine regulation of notch signaling. *Cancer Research* 72(15):3873–3885 DOI 10.1158/0008-5472.CAN-12-1060.



- Hu B, Thirtamara-Rajamani KK, Sim H, Viapiano MS. 2009.** Fibulin-3 is uniquely upregulated in malignant gliomas and promotes tumor cell motility and invasion. *Molecular Cancer Research* 7(11):1756–1770 DOI 10.1158/1541-7786.MCR-09-0207.
- Huang J, Sun Y, Chen H, Liao Y, Li S, Chen C, Yang Z. 2019.** ADAMTS5 acts as a tumor suppressor by inhibiting migration, invasion and angiogenesis in human gastric cancer. *Gastric Cancer* 22(2):287–301 DOI 10.1007/s10120-018-0866-2.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003.** Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264 DOI 10.1093/biostatistics/4.2.249.
- Ito-Kureha T, Koshikawa N, Yamamoto M, Semba K, Yamaguchi N, Yamamoto T, Seiki M, Inoue J-I. 2015.** Tropomodulin 1 expression driven by NF- $\kappa$ B enhances breast cancer growth. *Cancer Research* 75(1):62–72 DOI 10.1158/0008-5472.CAN-13-3455.
- Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T, Fukunaga R. 1998.** PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. *Molecular Cell* 2(6):877–885 DOI 10.1016/S1097-2765(00)80302-0.
- Kim IG, Kim SY, Choi SI, Lee JH, Kim KC, Cho EW. 2014.** Fibulin-3-mediated inhibition of epithelial-to-mesenchymal transition and self-renewal of ALDH+ lung cancer stem cells through IGF1R signaling. *Oncogene* 33(30):3908–3917 DOI 10.1038/onc.2013.373.
- Kim JS, Kurie JM, Ahn Y-H. 2015.** BMP4 depletion by miR-200 inhibits tumorigenesis and metastasis of lung adenocarcinoma cells. *Molecular Cancer* 14(1):173 DOI 10.1186/s12943-015-0441-y.
- Kim JI, Lakshmikanthan V, Frilot N, Daaka Y. 2010.** Prostaglandin E2 promotes lung cancer cell migration via EP4- $\beta$ Arrestin1-c-Src signalsome. *Molecular Cancer Research* 8(4):569–577 DOI 10.1158/1541-7786.MCR-09-0511.
- Kim HT, Yin W, Jin YJ, Panza P, Gunawan F, Grohmann B, Buettner C, Sokol AM, Preussner J, Guenther S, Kostin S, Ruppert C, Bhagwat AM, Ma X, Graumann J, Looso M, Guenther A, Adelstein RS, Offermanns S, Stainier DYR. 2018.** Myh10 deficiency leads to defective extracellular matrix remodeling and pulmonary disease. *Nature Communications* 9(1):4600 DOI 10.1038/s41467-018-06833-7.
- Kontic M, Stojsic J, Jovanovic D, Bunjevacki V, Ognjanovic S, Kuriger J, Puumala S, Nelson HH. 2012.** Aberrant promoter methylation of CDH13 and MGMT genes is associated with clinicopathologic characteristics of primary non-small-cell lung carcinoma. *Clinical Lung Cancer* 13(4):297–303 DOI 10.1016/j.clcc.2011.11.003.
- Kopparam J, Chiffelle J, Angelino P, Piersigilli A, Zangger N, Delorenzi M, Meylan E. 2017.** RIP4 inhibits STAT3 signaling to sustain lung adenocarcinoma differentiation. *Cell Death & Differentiation* 24(10):1761–1771 DOI 10.1038/cdd.2017.81.
- Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, Varella-Garcia M, Franklin WA, Aronson SL, Su PF, Shyr Y, Camidge DR, Sequist LV, Glisson BS, Khuri FR, Garon EB, Pao W, Rudin C, Schiller J, Haura EB, Socinski M, Shirai K, Chen H, Giaccone G, Ladanyi M, Kugler K, Minna JD, Bunn PA. 2014.** Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* 311(19):1998–2006 DOI 10.1001/jama.2014.3741.
- Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, Consonni D, Bertazzi PA, Wacholder S, Shih JH, Caporaso NE, Jen J. 2008.** Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLOS ONE* 3(2):e1651 DOI 10.1371/journal.pone.0001651.

- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6):882–883 DOI 10.1093/bioinformatics/bts034.
- Li T, Wang X, Jing L, Li Y. 2019. MiR-1-3p inhibits lung adenocarcinoma cell tumorigenesis via targeting protein regulator of cytokinesis 1. *Frontiers in Oncology* 9:120 DOI 10.3389/fonc.2019.00120.
- Li C, Xiong Y, Yang X, Wang L, Zhang S, Dai N, Li M, Ren T, Yang Y, Zhou SF, Gan L, Wang D. 2015. Lost expression of ADAMTS5 protein associates with progression and poor prognosis of hepatocellular carcinoma. *Drug Design, Development and Therapy* 9:1773–1783 DOI 10.2147/DDDT.S77069.
- Liang J, Wang J, Luo G, Pan Y, Wang X, Guo C, Zhang D, Yin F, Zhang X, Liu J, Wang J, Guo X, Wu K, Fan D. 2009. Function of PrPC (1-OPRD) in biological activities of gastric cancer cell lines. *Journal of Cellular and Molecular Medicine* 13(11–12):4453–4464 DOI 10.1111/j.1582-4934.2009.00687.x.
- Lin K, He S, He L, Chen J, Cheng X, Zhang G, Zhu B. 2014. Complement component 3 is a prognostic factor of non-small cell lung cancer. *Molecular Medicine Reports* 10(2):811–817 DOI 10.3892/mmr.2014.2230.
- Liu X, Li Y, Meng L, Liu XY, Peng A, Chen Y, Liu C, Chen H, Sun S, Miao X, Zhang Y, Zheng L, Huang K. 2018. Reducing protein regulator of cytokinesis 1 as a prospective therapy for hepatocellular carcinoma. *Cell Death & Disease* 9(5):534 DOI 10.1038/s41419-018-0555-4.
- Ma X, Kundu N, Rifat S, Walser T, Fulton AM. 2006. Prostaglandin E receptor EP4 antagonism inhibits breast cancer metastasis. *Cancer Research* 66(6):2923–2927 DOI 10.1158/0008-5472.CAN-05-4348.
- Maemura K, Watanabe K, Ando T, Hiyama N, Sakatani T, Amano Y, Kage H, Nakajima J, Yatomi Y, Nagase T, Takai D. 2018. Altered editing level of microRNAs is a potential biomarker in lung adenocarcinoma. *Cancer Science* 109(10):3326–3335 DOI 10.1111/cas.13742.
- Masui T, Hosotani R, Tsuji S, Miyamoto Y, Yasuda S, Ida J, Nakajima S, Kawaguchi M, Kobayashi H, Koizumi M, Toyoda E, Tulachan S, Arii S, Doi R, Imamura M. 2001. Expression of METH-1 and METH-2 in pancreatic cancer. *Clinical Cancer Research* 7:3437–3443.
- Mehan MR, Ayers D, Thirstrup D, Xiong W, Ostroff RM, Brody EN, Walker JJ, Gold L, Jarvis TC, Janjic N, Baird GS, Wilcox SK. 2012. Protein signature of lung cancer tissues. *PLOS ONE* 7(4):e35157 DOI 10.1371/journal.pone.0035157.
- Mehrpour M, Codogno P. 2010. Prion protein: from physiology to cancer biology. *Cancer Letters* 290(1):1–23 DOI 10.1016/j.canlet.2009.07.009.
- Nomoto S, Kanda M, Okamura Y, Nishikawa Y, Qiyong L, Fujii T, Sugimoto H, Takeda S, Nakao A. 2010. Epidermal growth factor-containing fibulin-like extracellular matrix protein 1, EFEMP1, a novel tumor-suppressor gene detected in hepatocellular carcinoma using double combination array analysis. *Annals of Surgical Oncology* 17(3):923–932 DOI 10.1245/s10434-009-0790-0.
- Pan Y, Zhao L, Liang J, Liu J, Shi Y, Liu N, Zhang G, Jin H, Gao J, Xie H, Wang J, Liu Z, Fan D. 2006. Cellular prion protein promotes invasion and metastasis of gastric cancer. *FASEB Journal* 20(11):1886–1888 DOI 10.1096/fj.06-6138fje.
- Park E-M, Lim Y-S, Ahn B-Y, Hwang SB. 2015. AAM-B interacts with nonstructural 4B and regulates hepatitis C virus propagation. *PLOS ONE* 10(7):e0132839 DOI 10.1371/journal.pone.0132839.

- Peng Y. 2006.** A novel ensemble machine learning for robust microarray data classification. *Computers in Biology and Medicine* **36(6)**:553–573 DOI [10.1016/j.combiomed.2005.04.001](https://doi.org/10.1016/j.combiomed.2005.04.001).
- Porter S, Scott SD, Sassoon EM, Williams MR, Jones JL, Girling AC, Ball RY, Edwards DR. 2004.** Dysregulated expression of adamalysin-thrombospondin genes in human breast carcinoma. *Clinical Cancer Research* **10(7)**:2429–2440 DOI [10.1158/1078-0432.CCR-0398-3](https://doi.org/10.1158/1078-0432.CCR-0398-3).
- Porter S, Span PN, Sweep FC, Tjan-Heijnen VC, Pennington CJ, Pedersen TX, Johnsen M, Lund LR, Romer J, Edwards DR. 2006.** ADAMTS8 and ADAMTS15 expression predicts survival in human breast carcinoma. *International Journal of Cancer* **118(5)**:1241–1247 DOI [10.1002/ijc.21476](https://doi.org/10.1002/ijc.21476).
- Qin L, Chen C, Chen L, Xue R, Ou-Yang M, Zhou C, Zhao S, He Z, Xia Y, He J, Liu P, Zhong N, Chen X. 2017.** Worldwide malaria incidence and cancer mortality are inversely associated. *Infectious Agents and Cancer* **12(1)**:14 DOI [10.1186/s13027-017-0117-x](https://doi.org/10.1186/s13027-017-0117-x).
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. 2007.** Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9(2)**:166–180 DOI [10.1593/neo.07112](https://doi.org/10.1593/neo.07112).
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015.** limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43(7)**:e47 DOI [10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007).
- Rodriguez-Rodero S, Fernandez AF, Fernandez-Morera JL, Castro-Santos P, Bayon GF, Ferrero C, Urdinguio RG, Gonzalez-Marquez R, Suarez C, Fernandez-Vega I, Fresno Forcelledo MF, Martinez-Cambor P, Mancikova V, Castelblanco E, Perez M, Marron PI, Mendiola M, Hardisson D, Santisteban P, Riesco-Eizaguirre G, Matias-Guiu X, Carnero A, Robledo M, Delgado-Alvarez E, Menendez-Torre E, Fraga MF. 2013.** DNA methylation signatures identify biologically distinct thyroid cancer subtypes. *Journal of Clinical Endocrinology & Metabolism* **98(7)**:2811–2821 DOI [10.1210/jc.2012-3566](https://doi.org/10.1210/jc.2012-3566).
- Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF, Laird-Offringa IA. 2012.** Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Research* **22**:1197–1211 DOI [10.1101/gr.132662.111](https://doi.org/10.1101/gr.132662.111).
- Shimo A, Nishidate T, Ohta T, Fukuda M, Nakamura Y, Katagiri T. 2007.** Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells. *Cancer Science* **98(2)**:174–181 DOI [10.1111/j.1349-7006.2006.00381.x](https://doi.org/10.1111/j.1349-7006.2006.00381.x).
- Su LJ, Chang CW, Wu YC, Chen KC, Lin CJ, Liang SC, Lin CH, Whang-Peng J, Hsu SL, Chen CH, Huang CY. 2007.** Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* **8**:140 DOI [10.1186/1471-2164-8-140](https://doi.org/10.1186/1471-2164-8-140).
- Sun KX, Zheng RS, Zeng HM, Zhang SW, Zou XN, Gu XY, Xia CF, Yang ZX, Li H, Chen WQ, He J. 2018.** The incidence and mortality of lung cancer in China, 2014. *Zhonghua Zhong Liu Za Zhi* **40**:805–811.
- Suzuki T, Kasamatsu A, Miyamoto I, Saito T, Higo M, Endo-Sakamoto Y, Shiiba M, Tanzawa H, Uzawa K. 2016.** Overexpression of TMOD1 is associated with enhanced regional lymph node metastasis in human oral cancer. *International Journal of Oncology* **48(2)**:607–612 DOI [10.3892/ijo.2015.3305](https://doi.org/10.3892/ijo.2015.3305).
- Toyooka S, Tokumo M, Shigematsu H, Matsuo K, Asano H, Tomii K, Ichihara S, Suzuki M, Aoe M, Date H, Gazdar AF, Shimizu N. 2006.** Mutational and epigenetic evidence for

- independent pathways for lung adenocarcinomas arising in smokers and never smokers. *Cancer Research* **66**(3):1371–1375 DOI [10.1158/0008-5472.CAN-05-2625](https://doi.org/10.1158/0008-5472.CAN-05-2625).
- Vargas AJ, Harris CC. 2016. Biomarker development in the precision medicine era: lung cancer as a case study. *Nature Reviews Cancer* **16**(8):525–537 DOI [10.1038/nrc.2016.56](https://doi.org/10.1038/nrc.2016.56).
- Vazquez F, Hastings G, Ortega MA, Lane TF, Oikemus S, Lombardo M, Iruela-Arispe ML. 1999. METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity. *Journal of Biological Chemistry* **274**(33):23349–23357 DOI [10.1074/jbc.274.33.23349](https://doi.org/10.1074/jbc.274.33.23349).
- Vencken SF, Greene CM, McKiernan PJ. 2015. Non-coding RNA as lung disease biomarkers: figure 1. *Thorax* **70**(5):501–503 DOI [10.1136/thoraxjnl-2014-206193](https://doi.org/10.1136/thoraxjnl-2014-206193).
- Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, Bergstrom S, Hanna L, Jakobsen E, Kolbeck K, Sundstrom S, Engholm G, Gavin A, Gjerstorff ML, Hatcher J, Johannesen TB, Linklater KM, McGahan CE, Steward J, Tracey E, Turner D, Richards MA, Rachet B. 2013. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007. *Thorax* **68**(6):551–564 DOI [10.1136/thoraxjnl-2012-202297](https://doi.org/10.1136/thoraxjnl-2012-202297).
- Wang Z, Cao CJ, Huang LL, Ke ZF, Luo CJ, Lin ZW, Wang F, Zhang YQ, Wang LT. 2015. EFEMP1 promotes the migration and invasion of osteosarcoma via MMP-2 with induction by AEG-1 via NF- $\kappa$ B signaling pathway. *Oncotarget* **6**(16):14191–14208 DOI [10.18632/oncotarget.3691](https://doi.org/10.18632/oncotarget.3691).
- Wei MM, Zhou GB. 2016. Long non-coding RNAs and their roles in non-small-cell lung cancer. *Genomics, Proteomics & Bioinformatics* **14**(5):280–288 DOI [10.1016/j.gpb.2016.03.007](https://doi.org/10.1016/j.gpb.2016.03.007).
- Xia Y, Yeddula N, Leblanc M, Ke E, Zhang Y, Oldfield E, Shaw RJ, Verma IM. 2012. Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model. *Nature Cell Biology* **14**(3):257–265 DOI [10.1038/ncb2428](https://doi.org/10.1038/ncb2428).
- Xin X, Majumder M, Girish GV, Mohindra V, Maruyama T, Lala PK. 2012. Targeting COX-2 and EP4 to control tumor growth, angiogenesis, lymphangiogenesis and metastasis to the lungs and lymph nodes in a breast cancer model. *Laboratory Investigation* **92**(8):1115–1128 DOI [10.1038/labinvest.2012.90](https://doi.org/10.1038/labinvest.2012.90).
- Yang T, Qiu H, Bao W, Li B, Lu C, Du G, Luo X, Wang L, Wan X. 2013. Epigenetic inactivation of EFEMP1 is associated with tumor suppressive function in endometrial carcinoma. *PLOS ONE* **8**(6):e67458 DOI [10.1371/journal.pone.0067458](https://doi.org/10.1371/journal.pone.0067458).
- Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16**(5):284–287 DOI [10.1089/omi.2011.0118](https://doi.org/10.1089/omi.2011.0118).
- Yue W, Dacic S, Sun Q, Landreneau R, Guo M, Zhou W, Siegfried JM, Yu J, Zhang L. 2007. Frequent inactivation of RAMP2, EFEMP1 and Dutt1 in lung cancer by promoter hypermethylation. *Clinical Cancer Research* **13**(15):4336–4344 DOI [10.1158/1078-0432.CCR-07-0015](https://doi.org/10.1158/1078-0432.CCR-07-0015).
- Zhai X, Li S-J. 2014. Methylation of RASSF1A and CDH13 genes in individualized chemotherapy for patients with non-small cell lung cancer. *Asian Pacific Journal of Cancer Prevention* **15**(12):4925–4928 DOI [10.7314/APJCP.2014.15.12.4925](https://doi.org/10.7314/APJCP.2014.15.12.4925).
- Zhan P, Xi G-M, Liu H-B, Liu Y-F, Xu W-J, Zhu Q, Zhou Z-J, Miao Y-Y, Wang X-X, Jin J-J, Lv T-F, Song Y. 2017a. Protein regulator of cytokinesis-1 expression: prognostic value in lung squamous cell carcinoma patients. *Journal of Thoracic Disease* **9**(7):2054–2060 DOI [10.21037/jtd.2017.06.91](https://doi.org/10.21037/jtd.2017.06.91).

- Zhan P, Zhang B, Xi G-M, Wu Y, Liu H-B, Liu Y-F, Xu W-J, Zhu Q-Q, Cai F, Zhou Z-J, Miu Y-Y, Wang X-X, Jin J-J, Li Q, Qian L-P, Lv T-F, Song Y. 2017b.** PRC1 contributes to tumorigenesis of lung adenocarcinoma in association with the Wnt/ $\beta$ -catenin signaling pathway. *Molecular Cancer* **16**(1):108 DOI [10.1186/s12943-017-0682-z](https://doi.org/10.1186/s12943-017-0682-z).
- Zhao P, Wu J, Lu F, Peng X, Liu C, Zhou N, Ying M. 2019.** The imbalance in the complement system and its possible physiological mechanisms in patients with lung cancer. *BMC Cancer* **19**(1):201 DOI [10.1186/s12885-019-5422-x](https://doi.org/10.1186/s12885-019-5422-x).
- Zhao X, Yang C, Wu J, Nan Y. 2018.** ADAMTS8 targets ERK to suppress cell proliferation, invasion, and metastasis of hepatocellular carcinoma. *OncoTargets and Therapy* **11**:7569–7578 DOI [10.2147/OTT.S173360](https://doi.org/10.2147/OTT.S173360).
- Zhong Y-H, Peng H, Cheng H-Z, Wang P. 2015.** Quantitative assessment of the diagnostic role of CDH13 promoter methylation in lung cancer. *Asian Pacific Journal of Cancer Prevention* **16**(3):1139–1143 DOI [10.7314/APJCP.2015.16.3.1139](https://doi.org/10.7314/APJCP.2015.16.3.1139).
- Zhu XJ, Liu J, Xu XY, Zhang CD, Dai DQ. 2014.** Novel tumor-suppressor gene epidermal growth factor-containing fibulin-like extracellular matrix protein 1 is epigenetically silenced and associated with invasion and metastasis in human gastric cancer. *Molecular Medicine Reports* **9**(6):2283–2292 DOI [10.3892/mmr.2014.2135](https://doi.org/10.3892/mmr.2014.2135).