Identification of Hub Genes Associated with Development of Lung Adenocarcinoma by Integrated Bioinformatics Analysis

Jinglei Li

China Academy of Chinese Medical Sciences Guanganmen Hospital https://orcid.org/0000-0003-0261-7891

Wei Hou (houwei64@126.com)

China Academy of Chinese Medical Sciences Guanganmen Hospital

Research

Keywords: lung adenocarcinoma, the cancer genome atlas, gene expression omnibus, weighted gene co-expression network analysis, Bioinformatics

Posted Date: June 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-550498/v1

Abstract

Purpose: Lung adenocarcinoma (LUAD) has high heterogeneity and poor prognosis, posing a major challenge to human health worldwide. Therefore, it is necessary to improve our understanding of the molecular mechanism of LUAD in order to be able to better predict its prognosis and develop new therapeutic strategies for target genes.

Methods: The Cancer Genome Atlas and Gene Expression Omnibus, were selected to comprehensively analyze and explore the differences between LUAD tumors and adjacent normal tissues. Critical gene information was obtained through weighted gene co-expression network analysis (WGCNA), differential gene expression analysis, and survival analysis.

Results: Using WGCNA and differential gene expression analysis, 29 differentially expressed genes were screened. The functional annotation analysis showed these genes to be mainly concentrated in heart trabecula formation, regulation of inflammatory response, collagen-containing extracellular matrix, and metalloendopeptidase inhibitor activity. Also, in the protein–protein interaction network analysis, 10 central genes were identified using Cytoscape's CytoHubba plug-in. The expression of *CDH5, TEK, TIMP3, EDNRB, EPAS1, MYL9, SPARCL1, KLF4*, and *TGFBR3* in LUAD tissue was found to be lower than that in the normal control group, while the expression of *MMP1* in LUAD tissue was higher than that in the normal control group. According to survival analysis, the low expression of *MYL9* and *SPARCL1* was correlated with poor overall survival in patients with LUAD. Finally, through the verification of the Oncomine database, it was found that the expression levels of *MYL9* and *SPARCL1* were consistent with the mRNA levels in LUAD samples, and both were downregulated.

Conclusion: Two survival-related genes, *MYL9* and *SPARCL1*, were determined to be highly correlated with the development of LUAD. Both may play an essential role in the development LUAD and may be potential biomarkers for its diagnosis and treatment in the future.

Introduction

The extremely high morbidity and mortality of lung cancer has become a significant health challenge of global concern [1]. It is estimated that lung cancer will account for 12% of the global cancers in 2020, and approximately one-fourth of the cancer deaths are caused by lung cancer [2], among which 80% of the cases exhibit the pathological type of non-small cell lung cancer (NSCLC) [3]. Lung adenocarcinoma (LUAD) is the most common tissue subtype of lung cancer [4]. Although surgery, chemotherapy, and radiotherapy have substantially improved patient prognosis, the patients' 5-year survival rate is only 15% [5]. Continued extensive research has led to the identification of biomarkers related to the prognosis of LUAD, such as epidermal growth factor receptor (EGFR). Although there are many generations of targeted drugs for lung cancer, such as gefitinib and alphatinib in the EGFR pathway, the possibility of drug resistance remains a challenge with important clinical implications [6]. With the development of next-

generation sequencing technology, rapid and advanced research can be undertaken to search for valuable biomarkers for the diagnosis, prognosis, and treatment of LUAD.

With the development of big bioinformatics data and the progress of genome sequencing technology, it is now possible to explore the mechanism of tumor occurrence and development in details. In systems biology, Weighted gene co-expression network analysis(WGCNA) approach uses unsupervised hierarchical clustering to effectively assign highly co-expressed genes into modules. Furthermore, it analyzes the relationship between the modules and clinical features, and thus enables critical genes in a module to be identified. After identification, the genes can be analyzed further based on their significance, as well as intra-modular connectivity [7]. Analysis of differential gene expression (DGE) is an essential part of transcriptomic analysis that offers an effective approach to the study of genomic regulation, as molecular mechanism. It also aids in revealing quantitative variations in levels of expression between the experimental and control groups. It plays an essential role in identifying potential biomarkers of specific diseases [8]. Therefore, to optimize the identification of genes that are highly related as candidate biomarkers, we combined WGCNA with DGE analysis.

We analyzed LUAD mRNA expression data in the Gene Expression Omnibus(GEO) as well as the The Cancer Genome Atlas(TCGA) web data resources using WGCNA and DGE analysis to obtain DEGs. We further explored the development mechanism of LUAD through Gene Ontology(GO) along with Kyoto Encyclopedia of Genes and Genomes(KEGG) enrichment analyses, protein–protein interaction(PPI) network analysis, and survival analysis, and verified our results using the Oncomine database. Herein, we aimed to analyze the differential co-expression of target genes in LUAD because it provides a potential basis for clinical diagnosis or therapy and improves our understanding of the potential molecular events underlying LUAD.

Material And Methods

Data processing and analysis of differential expression

We retrieved data on transcriptome RNA-seq of 551 LUAD cases, out of which 54 cases were non-tumor samples and 497 cases were tumor cases from the TCGA web data resource (https://portal.gdc.cancer.gov/) with level 3. Also, the normalized expression profiles of GSE8569 and GSE118370, the other two gene expression profiles of LUAD from GEO, were obtained by R packet GEOquery [9]. GSE8569 consists of 69 tumor samples as well as 6 matched non-tumor tissues from LUAD patients. GSE118370 consists of 6 tumor samples and 6 matched non-tumor tissues from LUAD patients. To find the DEGs in tumor samples versus the non-tumor ones, we employed statistics to analyze the TCGA-LUAD, GSE8569, and GSE118370 datasets with the Linear Models for Microarray Data (LIMMA) R package [10]. To adjust the P values for multiple tests, we used the Benjamini-Hochberg approach to control the false discovery rate (FDR). DEGs were those with the cut-off criteria of |log fold change (FC)| $\geq$ 1.0, and adjusted(adj.) $P < .05$. Meanwhile, the ggplot2 package helped us generate volcano maps of the DEGs.

Analysis of weighted gene co-expression network

WGCNA is effective in identifying gene sets that are highly synergistically altered. Also, it can be employed to select therapeutic targets and genes for candidate biomarker according to the how phenotype is related to gene set connectivity. Herein, we utilized the WGCNA [7] in R package to construct the profiles of TCGA-LUAD, GSE118370, and GSE8569 into gene co-expression networks based on gene expression data. To determine the key modules and establish a scale-free network, the soft-thresholding power was set to 2, 5, and 7, and scale-free R2 as > 0.85.Subsequently, we conducted Pearson's correlation test using the modules to analyze their relationship with the clinical features of LUAD. Statistical significance was set at adj. P < .05.

Identification of potential prognostic genes

To identify potential prognostic genes, we used Sangerbox Tools http://www.sangerbox.com/tool). Overlap the DEGs extracted from the co-expression network with TCGA-DEGs, GSE118370-DEGs, and GSE8569-DEGs

GO and KEGG pathway enrichment analyses the genes of interest

GO along with KEGG pathway enrichment analyses we conducted by utilizing the R package clusterProfiler [11], adj. p < .05 to further reveal the possible biological functions of intersecting DEGs. Besides, the biological processes, cell components, molecular function, as well as the KEGG pathways were visualized using the R package ggplot2 and GOplot.

Hub gene screening and PPI construction

Search Tool for the Retrieval of Interacting Genes(STRING ; https://string-db.org/) was utilized to generate a PPI network composed of selected genes. The tool is designed to predict PPIs. We chose the genes that had a score ≥ 0.15 to generate a network model visualized with Cytoscape (v3.8.0). We introduced the PPI network into the Cytoscape software platform and used the maximal clique centrality (MCC) function to comprehensively analyze the relationship between nodes, which was the most effective way to find hub genes in the co-expression network [12].

Survival analysis

To understand the association of the overall survival (OS) with the hub genes, Kaplan-Meier (KM) univariate survival assessment was conducted with the survival R package to analyze data on gene expression and survival (retrieved from the TCGA data resource). The association of the disease-free survival (DFS) with the expression of hub gene in LUAD patients was examined using the Gene Expression Profiling Interactive Analysis web server platform (http://gepia.cancerpku.cn/index.html). Herein, we only selected patients who did not dropout during follow-up for survival analysis. The patients were grouped based on the hub genes' median expression. The log-ranking p < .05 of the survival-related hub genes were considered statistically significant.

Verification of hub gene expression related to survival in the Oncomine database

The Oncomine web platform (https://www.oncomine.org/) was employed to explore the transcript levels of hub genes in LUAD versus non-tumor tissues following hub gene selection. The present data on **Oncomine consists of** 65 gene expression datasets containing about 48 million results on gene expression coming from more than 4700 microarray studies [13]. It has made a significant contribution to the in-depth analysis of differential genes.

Results

Generation of Weighted Gene Co-expression Modules

For each of the three datasets (TCGA-LUAD, GSE8569, GSE118370), we performed a series of steps for constructing a gene co-expression network, as outlined in, using the WGCNA software package available as an R package. With each module assigned a color, this study identified 13 modules in TCGA-LUAD, six modules in GSE8569 (excluding gray modules that are not assigned to any cluster), and 74 modules in GSE118370(Fig. 1). Then, we drew a heat map of three module-feature relationships to evaluate the association between each module and the two clinical features (normal and tumor). The turquoise module in TCGA-LUAD, the turquoise module in GSE8569, and turquoise module in GSE118370 had the highest correlation with normal tissues (TCGA-turquoise:r = 0.8,p = 9E − 122;GSE8569-turquoise:r = 0.62,p = 4E − 09;GSE118370-turquoise: r = 0.91,p = 4E − 05)(Fig. 2).

Identification of potential prognostic genes

A total of 3582 DEGs, 367 DEGs, and 789 DEGs were misexpressed in tumor tissues in TCGA, GSE8569, and GSE118370 datasets, respectively (|logFC|≥1.0, adj. P < 0.05). 7652, 450, and 1867 co-expressed genes were found in TCGA-turquoise module, GSE8569-turquoise module and GSE118370-turquoise module, respectively. Finally, 29 overlapping genes were extracted and used for follow-up analysis. (Fig. 3)

GO, KEGG pathway enrichment analyses

To comprehend further the potential function of 29 overlapping genes, we conducted the GO as well as the KEGG pathway enrichment assessment via the clusterProfiler software package. GO analysis showed that the biological process (BP) of 29 genes was primarily concentrated in heart trabecula formation, regulation of inflammatory response, and fat cell differentiation. The cellular component (CC) analysis showed that these genes primarily participates in the collagen-containing extracellular matrix, cell-cell junction, and stress fiber. Besides, metalloendopeptidase inhibitor activity and cytokine receptor binding were found to be related to these 29 genes in the molecular function (MF) analysis (P < .05). A rich analysis of the KEGG pathway with P < .05 showed that 29 overlapping genes were related to the complex

biological behavior of LUAD, such as "Leukocyte transendothelial migration," "Relaxin signaling pathway," "Chemokine signaling pathway," and "PPAR signaling pathway." The most significantly rich GO terms and KEGG pathways indicate that 29 genes interact at the functional level. (Fig. 4)

Development of the PPI network and assessment of hub genes

The PPI network of the 29 overlapping target genes was constructed via the STRING web platform (Fig. 5A). The selected hub genes from the PPI network using the MCC algorithm and cytoHubba plugin are shown in Fig. 5B. Cadherin 5 (CDH5), tek receptor tyrosine kinase (TEK), timp metallopeptidase inhibitor3 (TIMP3), endothelin receptor type B (EDNRB), endothelial pas domain protein 1 (EPAS1), myosin light chain 9 (MYL9), matrix metallopeptidase 1 (MMP1), SPARC-like protein 1 (SPARCL1), kruppel like factor 4 (KLF4), as well as the transforming growth factor-beta receptor 3 (TGFBR3), were selected as hub genes. (Fig. 5)

Verification of prognostic value of Hub gene

Based on the clinical information of TCGA-LUAD and GEPIA2 databases, 10 hub genes were analyzed by OS and DFS to investigate the prognostic value of hub genes in patients with LUAD. Kaplan-Meier analysis(Fig. 6) showed that among the 10 hub genes, the low expression of MYL9 and SPARCL1 was significantly correlated with OS in LUAD patients ($P < 0.05$), while in DFS (Fig. 7), the expression level of hub genes was not significantly different from that in LUAD patients ($P < .05$).

Verification of Hub Gene expression pattern

Using the Oncomine database, we obtained 428 items and 349 items of differential expression of MYL9 and SPARCL1 in different tumor types, respectively. A total of 57 studies revealed that there was a substantial difference in the expression of MYL9 between tumor tissues and non-tumor tissues, of which MYL9 was upregulated in 19 items and downregulated in 38 items (Fig. 8A). It is worth noting that MYL9 expression is suppressed in lung cancer (n = 11). A total of 74 studies revealed a considerable difference in the expression of SPARCL1 between tumor tissues and non-tumor tissues, including 23 upregulated and 51 downregulated items (Fig. 8B). As shown in Fig. 8B, the expression of SPARCL1 is downregulated in lung cancer (n = 11). There are 10 studies[14–23] in Oncomine about the differential MYL9 expression in lung cancer tissues and non-tumor tissues, including 1112 samples. The meta-analysis data revealed that the expression of MYL9 in lung cancer tissues was downregulated in contrast with that in non-tumor tissues (Median Rank = 333.0, P = 9.99E − 6) (Fig. 8C). Twelve studies [14–20, 22–26] analyzed expression difference of SPARCL1 between lung cancer tissues and non-tumor tissues, including 1236 samples. The meta-analysis data indicated that the expression of SPARCL1 in lung cancer tissues was downregulated in contrast with that in non-tumor tissues (Median Rank = 227.5, P = 6.76E − 9) (Fig. 8D). (Fig. 8)

Discussion

Among all categories of lung cancers, LUAD is one of the most common histological type. Determining the molecular mechanism underlying the occurrence and development of LUAD and the development of new therapeutic targets are essential means for improving its poor prognosis and low survival rate. Therefore, there is an urgent need to obtain comprehensive knowledge about the underlying molecular features of LUAD and identify more potential biomarkers for disease diagnosis and treatment.

In this study, we identified 29 significant genes that exhibited similar trend in their expression in TCGA, GSE8569, and GSE118370 databases by comprehensive bioinformatic analysis. Functional annotation assessment of the ClusterProfiler software package showed that the genes were primarily involved in heart trabecula formation and the regulation of inflammatory response, which are closely related to angiogenesis modulation, survival of endothelial cells, migration, proliferation, cell expansion and adhesion, recombination of the actin cytoskeleton, and maintenance of vascular resting. According to the MCC score of the CytoHubba plug-in, the first 10 genes related to LUAD were screened. It was found that the expression patterns of *CDH5, TEK, TIMP3, EDNRB, EPAS1, MYL9, SPARCL1, KLF4*, and *GFBR3* in LUAD tissues were lower in contrast with those in the non-tumor control group, while the expression pattern of *MMP1* in LUAD tissues was higher relative to the non-tumor control group. Among them, the low expression of *MYL9* and *SPARCL1* was found to be significantly related to patients' poor OS in LUAD. Lastly, the expression patterns of *MYL9* and *SPARCL1* were verified using the Oncomine database.

MYL9 is a protein-encoding gene [27]. The myosin regulatory subunit has an essential role in modulating smooth muscle as well as non-muscle cell contractile activity through its phosphorylation. Reflected in cytokinesis, cell locomotion, receptor capping [28, 29]. It is worth noting that myosin is also thought to participate in tumor progression as well as metastasis. *MYL9* facilitates airway smooth muscle cell migration and contraction, migration of megakaryocyte and tip cell under pathological conditions such as angiogenesis and tumor growth during development [30–32]. *MYL9* is modulated by the troponin-related transcription factor-serum response factor (MRTF-SRF) cascade and is necessary for the migration of tumor cells, megakaryocytes, and TIP cells in vivo. The upregulation of *MYL9* can also promote the invasive function of tumor-related fibroblasts. Our research shows that the leukocyte transendothelial migration signaling pathway is closely related to *MYL9*, an essential mechanism for transporting white blood cells to the site of injury, immune response, or infection in the inflammatory response [33]. Many researchers have found that *MYL9* is expressed in a variety of tumors. Some studies have shown that *MYL9* upregulation is remarkably linked to the shorter survival time of patients with colorectal cancer [34–36]. Studies by Kruthika et al. have shown that the high expression of *MYL9* in patients with glioblastoma is related to poor prognosis and plays an essential role in improving tumor invasiveness [37]. Huang et al. found that the downregulation of *MYL9* in stroma indicates progression of prostate cancer and poor biochemical recurrence-free survival [38]. Our results showed that a higher level of *MYL9* in tumor tissue was linked to good disease outcomes LUAD patients. We speculate that *MYL9* may be a human tumor suppressor gene.

Glycoprotein SPARCL1 belongs to the SPARC family of matricellular proteins [39]. GO analysis showed that the biological process closely related to *SPARCL1* was cell-cell adhesion through plasma-membrane adhesion molecules. Cell adhesion molecules are at least partly embedded in the cell membrane. It is related to the modulation of cell adhesion, proliferation, migration [39]. Many studies have shown that *SPARCL1* has different expression patterns in different types of tumors and plays the tumor suppressor and angiogenesis regulator [40–42]. Nelson et al. [43] demonstrated for the first time that the transcriptional level of *SPARCL1* is lower in transformed prostate epithelial cell lines and metastatic prostate cancer cells, strongly suggesting the hypothetical role of *SPARCL1* in inhibiting the progression and metastasis of prostate cancer. A comprehensive analysis by Wu et al. [44] showed that the downregulated expression of *SPARCL1* might be vital for the formation and pathogenesis of cervical cancer. It is also related to precancerous lesions and migration in the occurrence of cervical cancer. Ma et al. [45] found that *SPARCL1* protein expression in ovarian cancer tissues is lower in contrast with the neighboring non-tumor tissues, which may suppress the migration along with the proliferative ability of ovarian cancer cells by downregulating the *MEK/ERK* signaling pathway. The results of Han et al. [46] indicated that the expression of *SPARCL1* protein in colorectal tumors was substantially lower compared to corresponding non-tumor tissues, which may be used as a potential tumor suppressor gene and associated with a good prognosis. In our study, relative to non-tumor tissues, *SPARCL1* expression was lower in tumor tissues, which was near related to LUAD. The survival analysis data revealed that the higher level of *SPARCL1* in tumor tissue was linked to the good prognosis of patients with LUAD.

Although this study screened the DEGs of LUAD tumor tissue and neighboring non-tumor tissue from the publicly downloaded data sets and conducted a comprehensive bioinformatic analysis, the molecular mechanism of survival-related genes influencing the prognosis of LUAD patients should be further verified to validate our findings.

Conclusion

This study applied a comprehensive transcriptomic approach to analyze available transcriptomic data to identify key prognostic genes in LUAD. We identified the low expression of MYL9 and SPARCL1 was found to be significantly related to patients' poor OS in LUAD.These findings require validation by further interventional and functional studies.

Abbreviations

LUAD: Lung adenocarcinoma; WGCNA: weighted gene co-expression network analysis; NSCLC: non-small cell lung cancer ;DGE: differential gene expression; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein–protein interaction; LIMMA: Linear Models for Microarray Data; FC: log fold change; FDR: false discovery rate; STRING: Search Tool for the Retrieval of Interacting Genes; MCC : maximal clique centrality ; OS : overall survival ; KM : Kaplan-Meier ; DFS : disease-free survival ; BP: biological process; CC: cellular

component; MF: molecular function; CDH5: Cadherin 5; TEK: tek receptor tyrosine kinase; TIMP3: timp metallopeptidase inhibitor3;EDNRB :endothelin receptor type B; EPAS1: endothelial pas domain protein 1; MYL9: myosin light chain 9; MMP1 :matrix metallopeptidase 1; SPARCL1: SPARC-like protein 1; KLF4:kruppel like factor 4; TGFBR3: transforming growth factor-beta receptor 3.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

JL and WH designed the research study. JL took responsibility for statistical analyses. JL wrote the manuscript. WH evaluated and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was financially supported through grants from the National Natural Science Foundation of China (No. 82074239).

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its Supplementary information files].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' information

Jinglei Li [1,2],Wei Hou [1*]

1 Department of Oncology, Guang'anmen Hospital of China Academy of Chinese Medical Sciences,Beijing,100053,China

2 Beijing University of Chinese Medicine, Beijing, 100029, China

Correspondence: Wei Hou; Tel: +010-88001500; Fax: NA; Email: houwei64@126.com

First author: Jinglei Li; Tel :86-18810211813 ;Fax :NA; Email :lijinglei535@foxmail.com

References

1. Carlisle J, Steuer C, Owonikoko T, Saba N. An update on the immune landscape in lung and head and neck cancers. *CA Cancer J Clin*. 2020.
2. Siegel R, Miller K, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30.
3. Miller K, Nogueira L, Mariotto A, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin. 2019;69(5):363–85.
4. Inamura K. Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *Int J Mol Sci*. 2018;19(4).
5. Herbst R, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553(7689):446–54.
6. Wei C, Yao X, Jiang Z, et al. Cordycepin Inhibits Drug-resistance Non-small Cell Lung Cancer Progression by Activating AMPK Signaling Pathway. Pharmacol Res. 2019;144:79–89.
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
8. Segundo-Val I, Sanz-Lozano C. Introduction to the Gene Expression Analysis. Methods Mol Biol. 2016;1434:29–43.
9. Davis S, Meltzer P. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23(14):1846–7.
10. Ritchie M, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
11. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284–7.
12. Chin C, Chen S, Wu H, Ho C, Ko M, Lin C. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014:S11.
13. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* Jan-Feb. 2004;6(1):1–6.
14. Beer D, Kardia S, Huang C, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002;8(8):816–24.
15. Stearman RS, Dwyer-Nield L, Zerbe L, et al. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. *Am J Pathol* Dec. 2005;167(6):1763–75.

16. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* Nov. 2001;20(24):13790–5. 98(.
17. Su LJ, Chang CW, Wu YC, et al. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* Jun. 2007;1:8.
18. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* Nov. 2001;20(24):13784–9. 98(.
19. Landi MT, Dracheva T, Rotunno M, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE* Feb. 2008;20(2):e1651. 3(.
20. Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* Jan. 2012;1(1):100–11. 72(.
21. Selamat SA, Chung BS, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res* Jul. 2012;22(7):1197–211.
22. Hou J, Aerts J, den Hamer B, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS ONE* Apr. 2010;22(4):e10312. 5(.
23. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* Dec. 2005;1(23):4205–8. 21(.
24. Selamat S, Chung B, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res. 2012;22(7):1197–211.
25. Yamagata N, Shyr Y, Yanagisawa K, et al. A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res* Oct. 2003;15(13):4695–704. 9(.
26. Talbot SG, Estilo C, Maghami E, et al. Gene expression profiling allows distinction between primary and metastatic squamous cell carcinomas in the lung. *Cancer Res* Apr. 2005;15(8):3063–71. 65(.
27. Licht AH, Nubel T, Feldner A, et al. Junb regulates arterial contraction capacity, cellular contractility, and motility via its target Myl9 in mice. *J Clin Invest* Jul. 2010;120(7):2307–18.
28. Iwasaki T, Murata-Hori M, Ishitobi S, Hosoya H. Diphosphorylated MRLC is required for organization of stress fibers in interphase cells and the contractile ring in dividing cells. Cell Struct Funct. 2001;26(6):677–83.
29. Kumar C, Mohan S, Zavodny P, Narula S, Leibowitz P. Characterization and differential expression of human vascular smooth muscle myosin light chain 2 isoform in nonmuscle cells. Biochemistry. 1989;28(9):4027–35.
30. Xu W, Hong W, Shao Y, Ning Y, Cai Z, Li Q. Nogo-B regulates migration and contraction of airway smooth muscle cells by decreasing ARPC 2/3 and increasing MYL-9 expression. Respir Res. 2011;12:14. doi:10.1186/1465-9921-12-14.
31. Gilles L, Bluteau D, Boukour S, et al. MAL/SRF complex is involved in platelet formation and megakaryocyte migration by regulating MYL9 (MLC2) and MMP9. Blood. 2009;114(19):4221–32.

32. Franco C, Blanc J, Parlakian A, et al. SRF selectively controls tip cell invasive behavior in angiogenesis. Development. 2013;140(11):2321–33.
33. Luscinskas F, Ma S, Nusrat A, Parkos C, Shaw S. Leukocyte transendothelial migration: a junctional affair. Semin Immunol. 2002;14(2):105–13.
34. Qiu X, Cheng S, Xu F, Yin J, Wang L, Zhang X. Weighted gene co-expression network analysis identified MYL9 and CNN1 are associated with recurrence in colorectal cancer. J Cancer. 2020;11(8):2348–59.
35. Zhu K, Wang Y, Liu L, Li S, Yu W. Long non-coding RNA MBNL1-AS1 regulates proliferation, migration, and invasion of cancer stem cells in colon cancer by interacting with MYL9 via sponging microRNA-412-3p. Clin Res Hepatol Gastroenterol. 2020;44(1):101–14.
36. Zhou Y, Bian S, Zhou X, et al. Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. *Cancer Cell*. 2020.
37. Kruthika B, Sugur H, Nandaki K, Arimappamagan A, Paturu K, Santosh V. Expression pattern and prognostic significance of myosin light chain 9 (MYL9): a novel biomarker in glioblastoma. J Clin Pathol. 2019;72(10):677–81.
38. Huang Y, Han Z, Liang Y, et al. Decreased expression of myosin light chain MYL9 in stroma predicts malignant progression and poor biochemical recurrence-free survival in prostate cancer. *Medical oncology (Northwood, London, England)*. 2014;31(1):820. doi:10.1007/s12032-013-0820-4.
39. Gagliardi F, Narayanan A, Mortini P. SPARCL1 a novel player in cancer biology. Crit Rev Oncol Hematol. 2017;109:63–8.
40. Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* Jun. 2009;4(7247):717–21. 459(.
41. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. 2009;11(1):R7.
42. Skrzypczak M, Goryca K, Rubel T, et al. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS ONE*. Oct 1 2010;5(10).
43. Nelson P, Plymate S, Wang K, et al. Hevin, an antiadhesive extracellular matrix protein, is down-regulated in metastatic prostate adenocarcinoma. Cancer Res. 1998;58(2):232–6.
44. Wu D, Shi J, Liu T, Deng S, Han R, Xu Y. Integrated analysis reveals down-regulation of SPARCL1 is correlated with cervical cancer development and progression. Cancer Biomark A. 2018;21(2):355–65.
45. Ma Y, Xu Y, Li L. SPARCL1 suppresses the proliferation and migration of human ovarian cancer cells via the MEK/ERK signaling. Exp Ther Med. 2018;16(4):3195–201.
46. Han W, Cao F, Ding W, et al. Prognostic value of SPARCL1 in patients with colorectal cancer. Oncol Lett. 2018;15(2):1429–34.

Figures

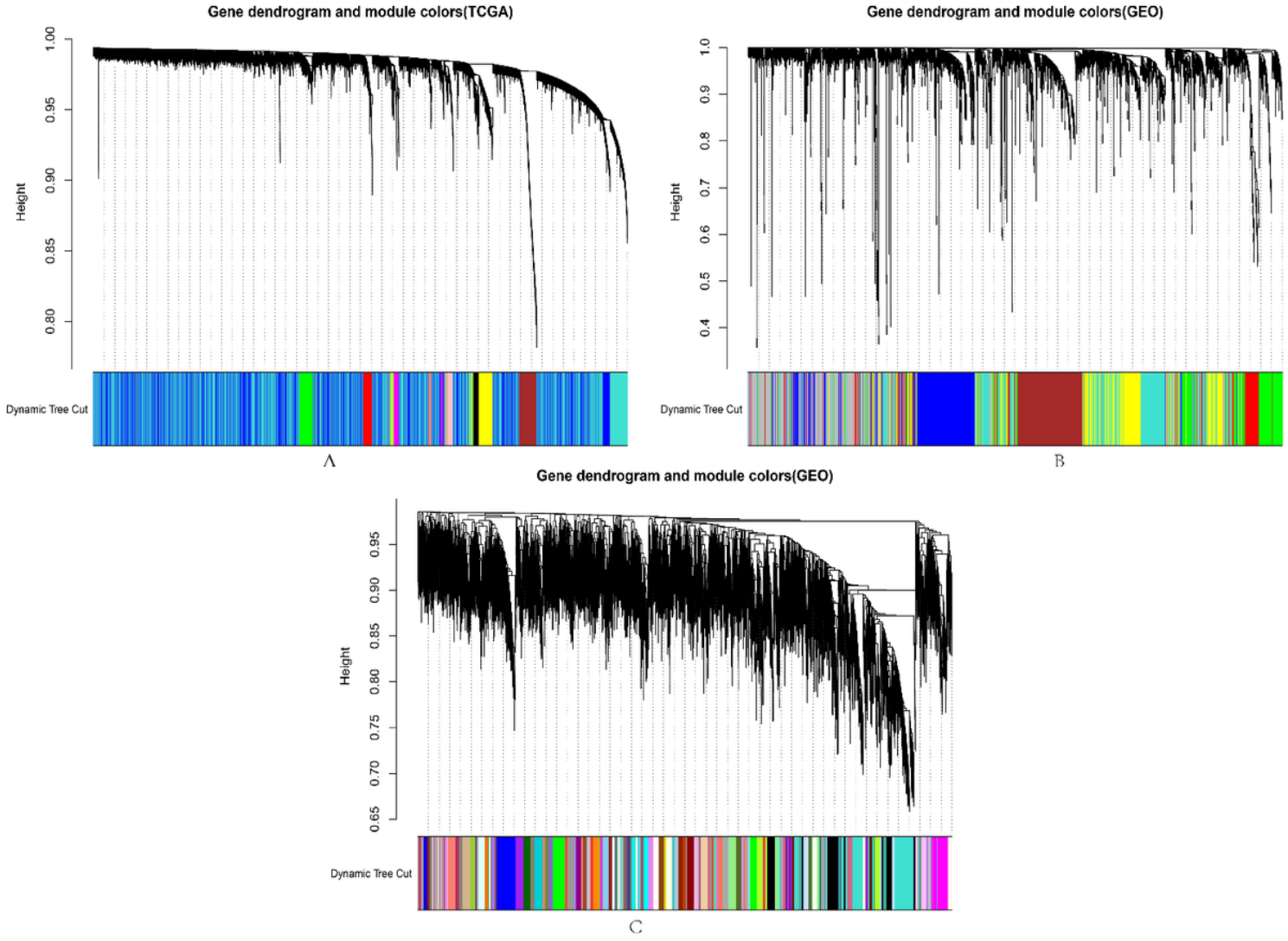

Figure 1

Identification of modules associated with the clinical information in TCGA-LUAD,GSE8569,GSE118370 datasets. The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors. (A)TCGA-LUAD,(B)GSE8569,(C)GSE118370.

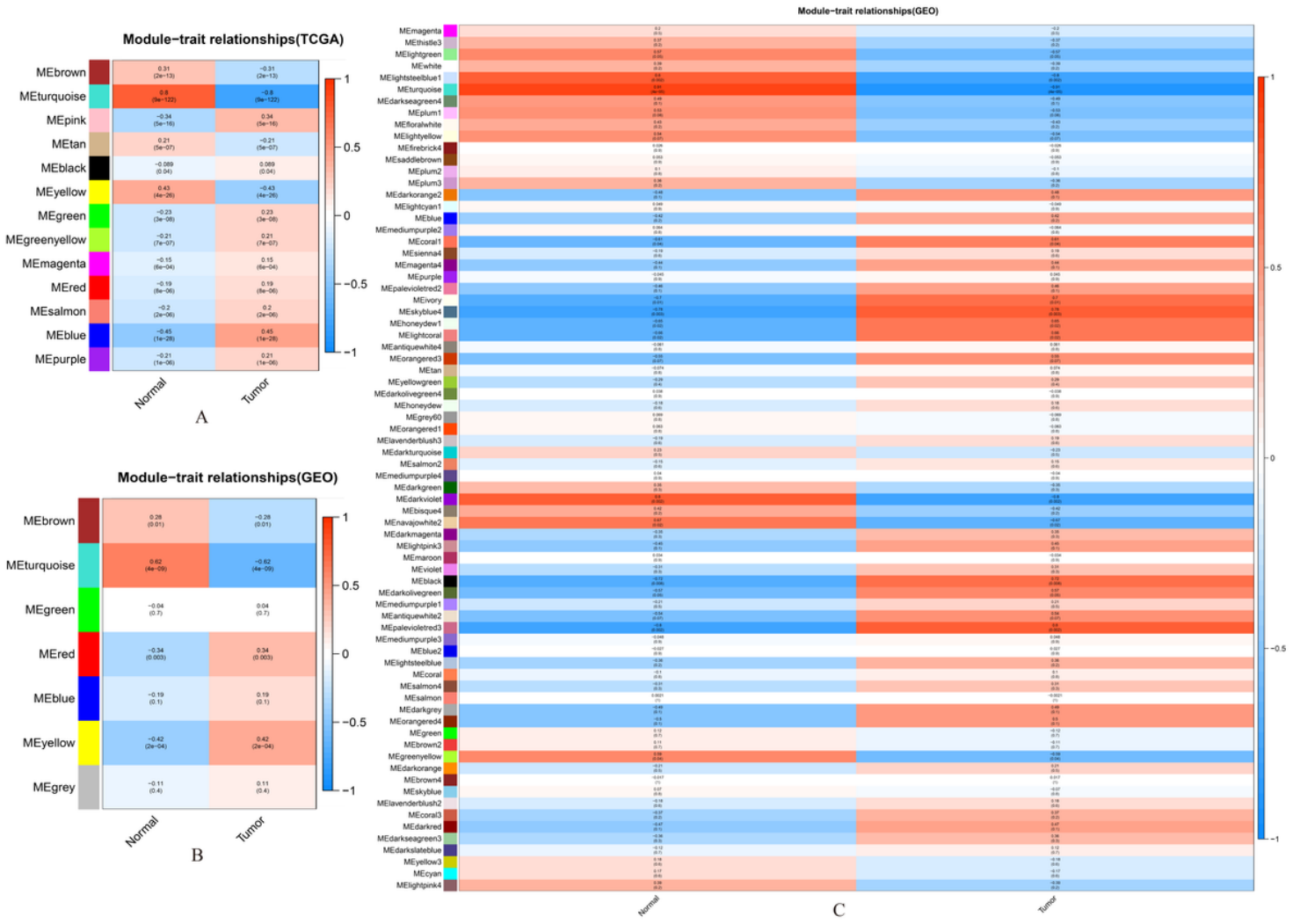

Figure 2

Module-trait associations. Every row represents a color module, whereas the column represents to a clinical parameter (cancer and non-tumor). Every cell contains the respective association and P-value. (A)TCGA-LUAD,(B)GSE8569,(C)GSE118370.

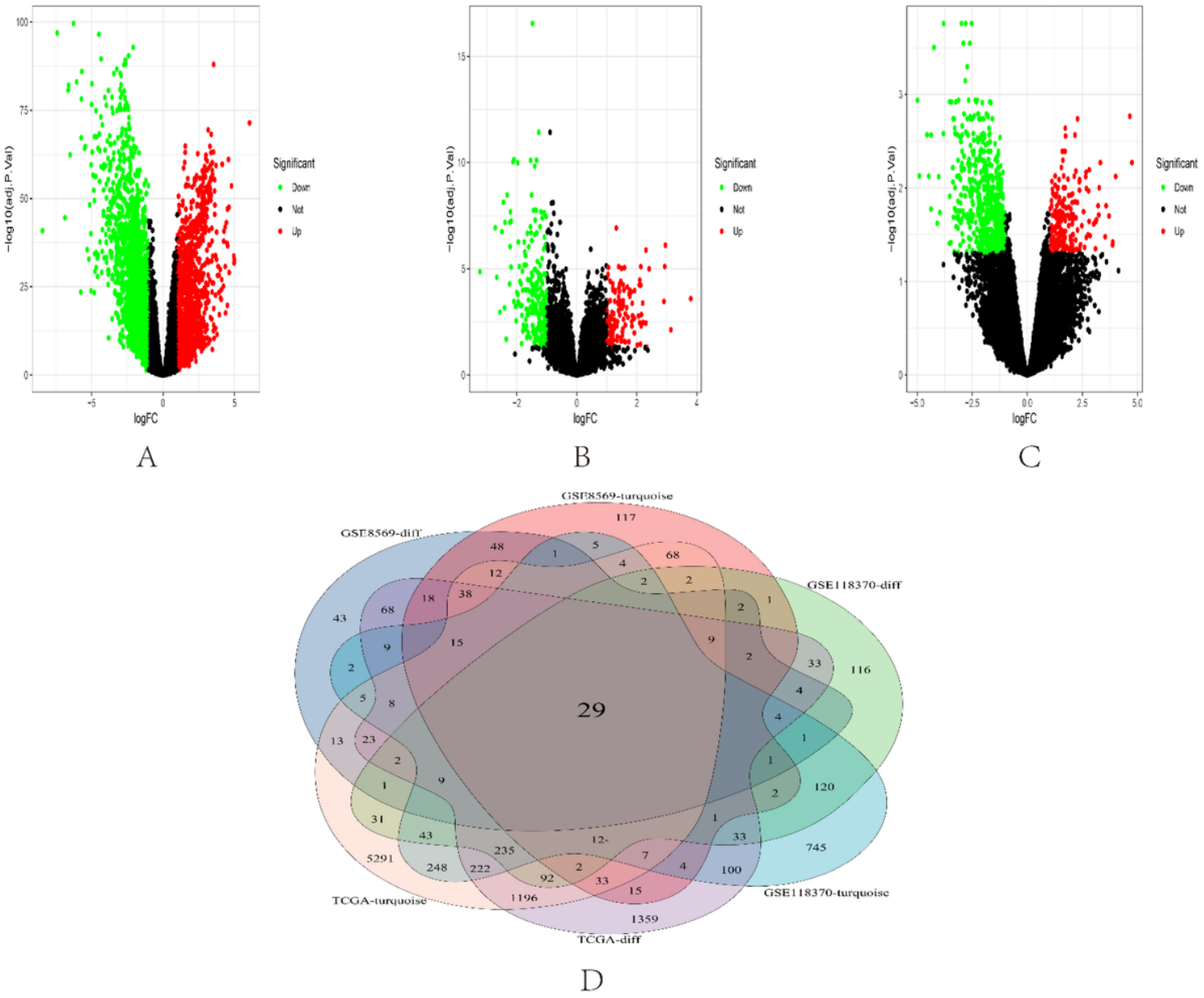

Figure 3

Determination of the DEGs between the TCGA, GSE8569 and GSE118370 cohorts of LUAD with the cut-off criteria of adj. P < .05 as well as |logFC| ≥ 1.0 (A) Volcano plot of DEGs in the TCGA dataset. (B) Volcano plot of DEGs in the GSE8569 dataset. (C) Volcano scatter plot of DEGs in the GSE118370 cohort. (D) The Venn diagram of genes among DEG lists and co-expression module. Overall, 29 overlapping genes in the intersection of DEG lists and three co-expression modules.

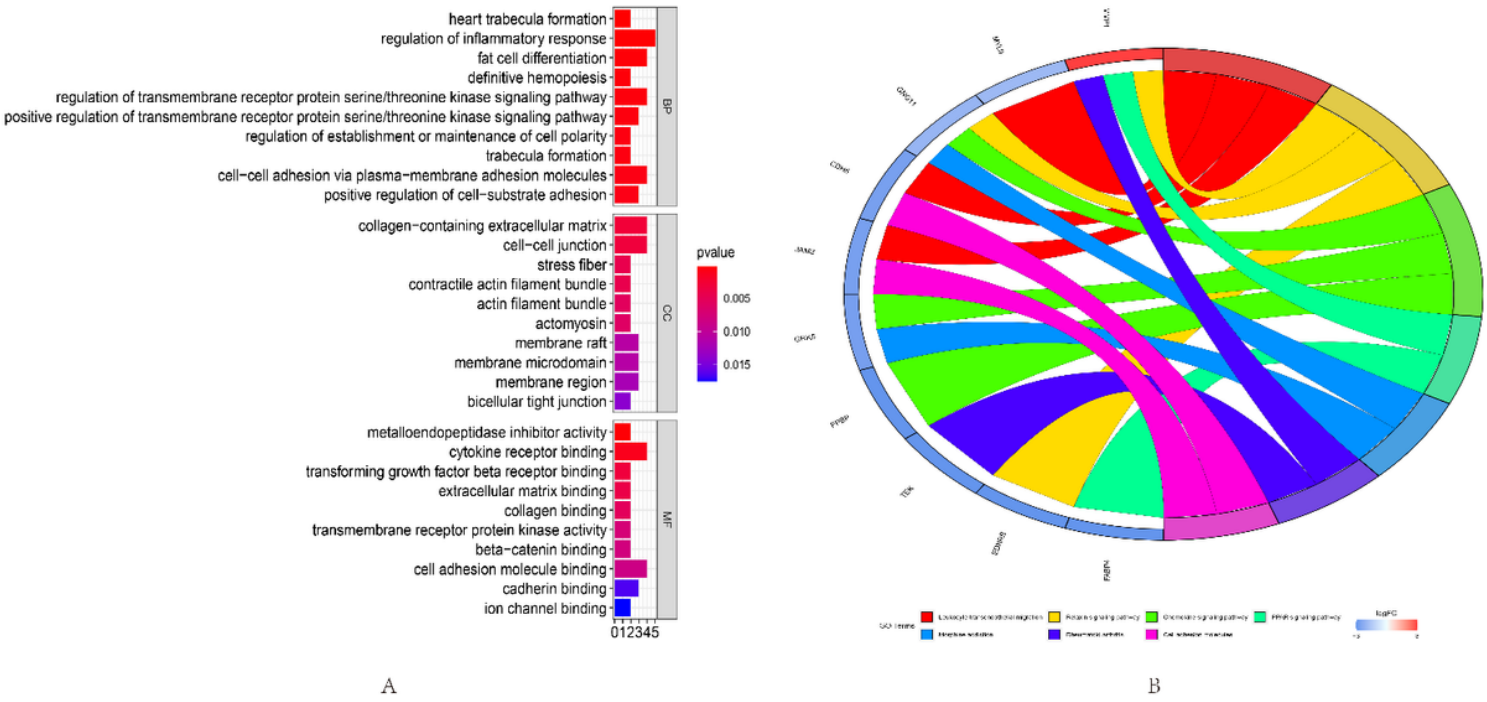

Figure 4

GO, KEGG pathway enrichment analyses.(A) GO enrichment analyses of 29 overlapping genes.(B) KEGG pathway enrichment analyses of 29 overlapping genes.

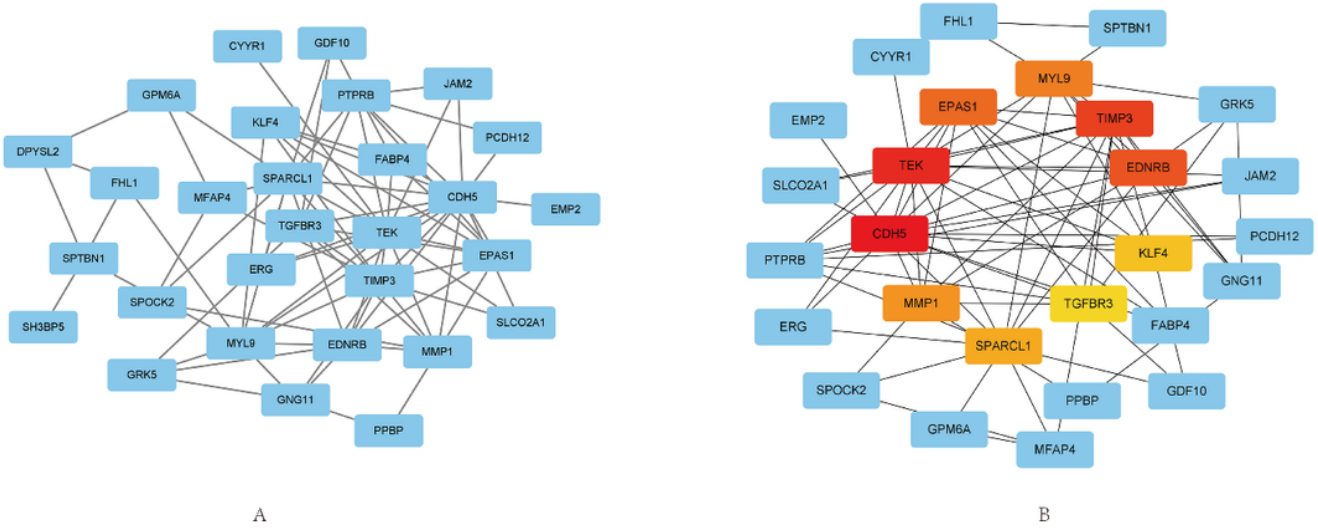

Figure 5

PPI network as well as the visualization of the candidate hub gene. (A) PPI network of the genes: DEG lists vs. three co-expression modules. The genes are represented by blue nodes. Interaction between the nodes are represented by edges. (B) Uncovering of the hub genes from the PPI network via the maximal clique centrality (MCC) algorithm. Edges indicates the protein-protein cross talks. High MCC sores genes are represented by red nodes, whereas, the yellow node indicates low MCC sore genes.

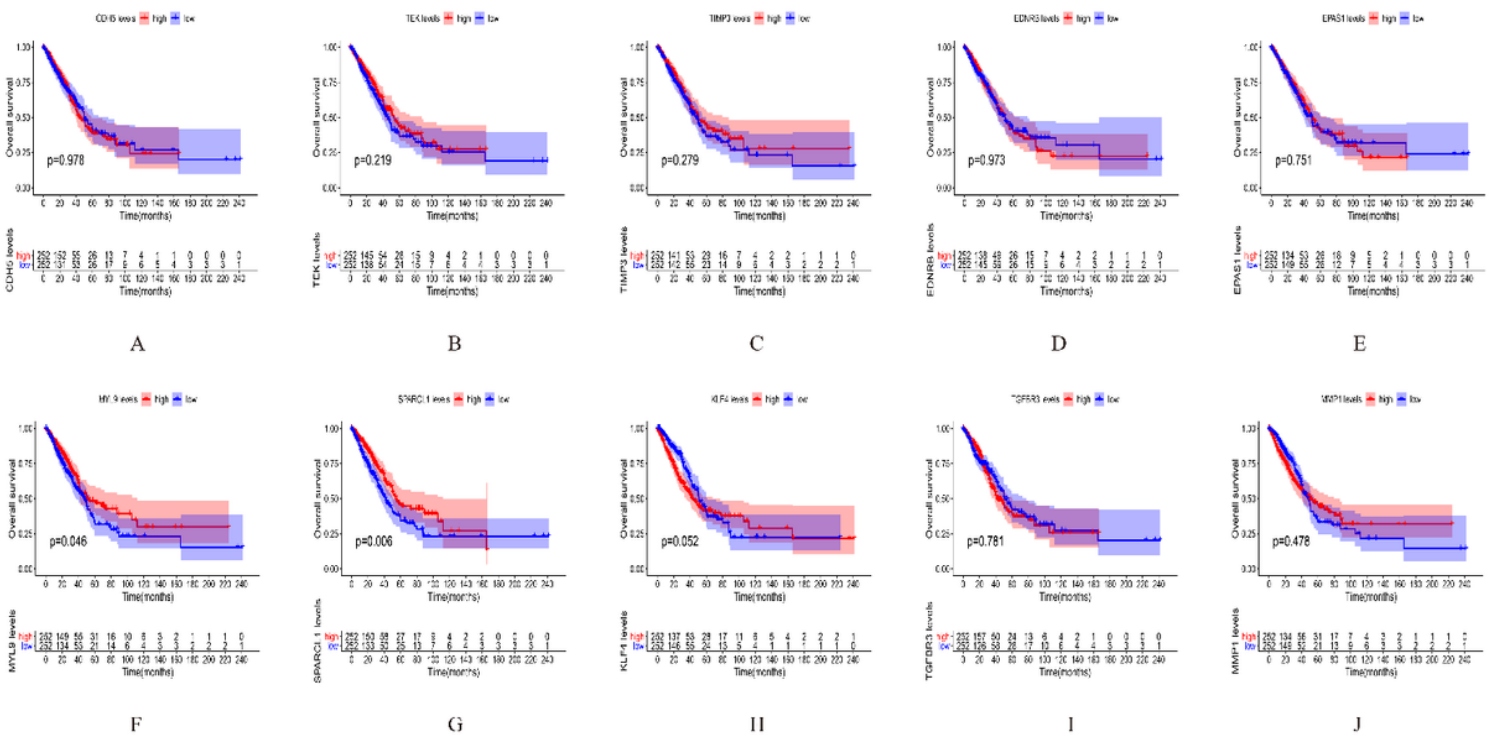

Figure 6

Overall survival (OS) evaluation of 10 hub genes in LUAD patients from the TCGA data resource. Survival assessment for: (A) CDH5 in LUAD. (B) TEK in LUAD. (C) TIMP3 in LUAD. (D) DNRB in LUAD. (E) EPAS1 in LUAD. (F) MYL9 in LUAD. (G) SPARCL1 in LUAD. (H) KLF4 in LUAD. (I) TGFBR3 in LUAD. (J) MMP1 in LUAD. Patients were classified into the low-level group (blue) and high-level group (red) based on median expression of the gene (Log-rank P <.05).

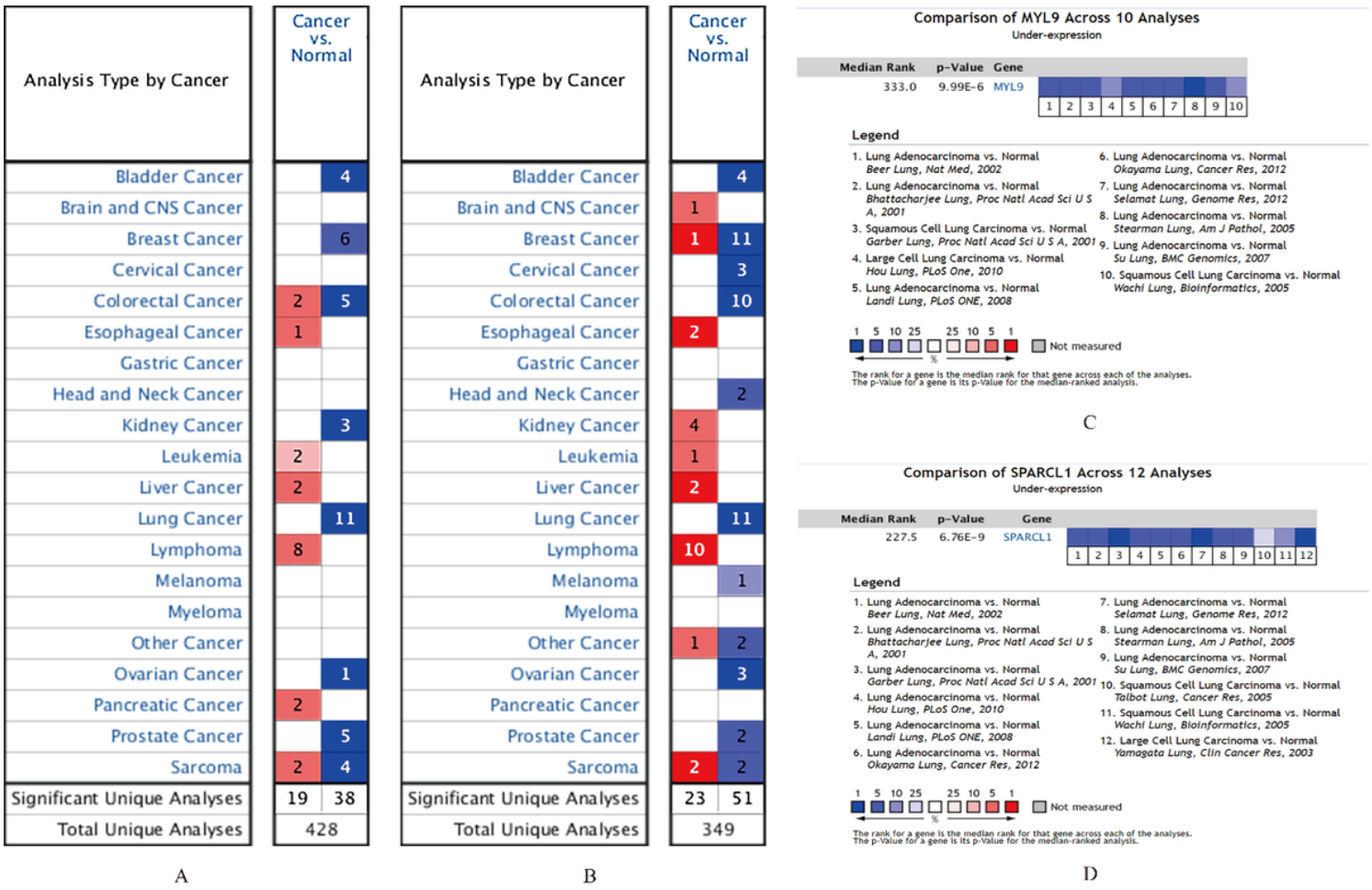

Figure 7

Differential expression of the hub gene in tumors and expression profile of hub genes in lung cancer tissues in the Oncomine data resource. (A). Differential expressions of MYL9 in tumors. (B). Differential expressions of SPARCL1 in tumors. (C). Expression profile of MYL9 in the lung cancer tissues according to the Oncomine data resource. (D). Expression profile of SPARCL1 lung cancer tissues according to the Oncomine data resource.