

# Bioinformatics Identification of Key Genes for the Development and Prognosis of Lung Adenocarcinoma

INQUIRY: The Journal of Health Care Organization, Provision, and Financing  
Volume 59: 1–10  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00469580221096259  
[journals.sagepub.com/home/inq](https://journals.sagepub.com/home/inq)  


Xuan Luo, PhD<sup>1,#</sup> , Jian Guo Xu, PhD<sup>2,#</sup> , ZhiYuan Wang, PhD<sup>1</sup> , XiaoFang Wang, PhD<sup>3</sup> , QianYing Zhu, master's degree<sup>1</sup> , Juan Zhao, master's degree<sup>1</sup> , and Li Bian, PhD<sup>1</sup> 

## Abstract

**Objective:** Lung adenocarcinoma (LUAD) is a common malignant tumor with a poor prognosis. The present study aimed to screen the key genes involved in LUAD development and prognosis. **Methods:** The transcriptome data for 515 LUAD and 347 normal samples were downloaded from The Cancer Genome Atlas and Genotype Tissue Expression databases. The weighted gene co-expression network and differentially expressed genes were used to identify the central regulatory genes for the development of LUAD. Univariate Cox, LASSO, and multivariate Cox regression analyses were utilized to identify prognosis-related genes. **Results:** The top 10 central regulatory genes of LUAD included IL6, PECAM1, CDH5, VWF, THBS1, CAV1, TEK, HGF, SPP1, and ENG. Genes that have an impact on survival included PECAM1, HGF, SPP1, and ENG. The favorable prognosis genes included KDF1, ZNF691, DNASE2B, and ELAPOR1, while unfavorable prognosis genes included RPL22, ENO1, PCSK9, SNX7, and LCE5A. The areas under the receiver operating characteristic curves of the risk score model in the training and testing datasets were .78 and .758, respectively. **Conclusion:** Bioinformatics methods were used to identify genes involved in the development and prognosis of LUAD, which will provide a basis for further research on the treatment and prognosis of LUAD.

## Keywords

lung adenocarcinoma, prognosis, weighted gene co-expression network analysis

### What Do We Already Know About This Topic?

Genes affecting the development of LUAD included PECAM1, HGF, SPP1, ENG, and The favorable prognosis genes included KDF1, ZNF691, DNASE2B, and ELAPOR1, while the unfavorable prognosis genes included RPL22, ENO1, PCSK9, SNX7, and LCE5A.

### How Does Your Research Contribute to the Field?

Bioinformatics methods were used to identify genes involved in the development and prognosis of LUAD, which will provide a basis for further research on its treatment and prognosis.

<sup>1</sup>The First Affiliated Hospital of Kunming Medical University, Kunming, China

<sup>2</sup>Department of Dental Research, The Affiliated Stomatological Hospital of Kunming Medical University, Kunming, China

<sup>3</sup>Department of Pathology, The Second Affiliated Hospital of Kunming Medical University, Kunming, China

#These authors contributed equally to this study.

### Corresponding Author:

Bian Li, Department of Pathology, The First Affiliated Hospital of Kunming Medical University, Kunming 650032, China.

Email: [1422183882@qq.com](mailto:1422183882@qq.com)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and

Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

### What Are Your Research's Implications Towards Theory, Practice, or Policy?

This study provides basic evidence for the diagnosis and treatment of LUAD.

## Background

Lung adenocarcinoma (LUAD) is a common malignant tumor with high morbidity and mortality.<sup>1</sup> Its 5-year survival rate is 4-17%.<sup>2</sup> LUAD treatment is difficult due to tumor heterogeneity, and the risk of recurrence after treatment is also higher.<sup>3</sup> With the development of genomics and use of bioinformatics analysis for lung cancer, a large number of molecular markers related to lung cancer development, drug resistance, and prognosis have been discovered, such as epidermal growth factor receptor tyrosine kinase inhibitors (EGFR TKIs) and anaplastic lymphoma kinase (ALK) inhibitors.<sup>4,5</sup> Therefore, it is particularly important to study the biological markers of LUAD development and prognosis. The weighted gene co-expression network analysis (WGCNA) is a better method to screen central regulatory genes for tumorigenesis and development.<sup>6</sup> WGCNA explores the relationship between genes and phenotypes by weighting the correlation network.<sup>7</sup> It can analyze signaling networks by converting gene expression data into co-expression modules.<sup>8</sup> It is also used to screen genes associated with cancer-related modules and signatures. WGCNA finds relevant genes and predicts gene functions by analyzing key genes and identifying potential therapeutic targets and predictive biomarkers.<sup>9</sup> Wei et al<sup>10</sup> have adopted WGCNA to analyze LUAD and have identified modules that were highly correlated with LUAD. Yi et al<sup>11</sup> have identified the occurrence- and prognosis-related genes of LUAD using WGCNA. In recent years, WGCNA has been predominantly used to screen important hub genes against LUAD gene expression data. The present study was based on the transcriptome data and clinical information on LUAD from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases. The development regulatory genes in LUAD were screened using WGCNA as well as differentially expressed genes (DEGs). Univariate Cox, LASSO, and multivariate Cox regression analyses were utilized to screen the key genes for the prognosis of LUAD and to construct a prognostic model in order to ensure further understanding of the developmental and prognostic LUAD markers. Experimental design (Figure 1).

## Materials and Methods

### Data Acquisition

Transcriptome data (mRNA) for LUAD were downloaded from the UCSC Xena database ([https://xenabrowser.net/](https://xenabrowser.net/datapages/)

[datapages/](https://xenabrowser.net/datapages/)) TCGA in Pan-Cancer (PANCAN). Normal sample transcriptome data for LUAD were downloaded from the UCSC Xena database (<https://xenabrowser.net/datapages/>) GTEx. The clinical data were downloaded from the TCGA. Final cohort included a total of 515 LUAD and 347 normal samples. Ethical approval was not required for the present study because the data were obtained from public databases.

### Intersection of Differentially Expressed Genes and Weighted Gene Co-expression Network Analysis

The R package limma was used to screen DEGs. Conditions for screening included log fold change (logFC) of  $\geq |2|$  and adjusted  $P \leq .05$ . R package WGCNA was used to analyze genes with TPM  $\geq 1$ .  $\beta$  values and scale-free  $R^2$  were adjusted as a soft-threshold index to construct a scale-free co-expression network. WGCNA was used to screen out the relevant LUAD modules. The most relevant LUAD modules and genes intersecting with DEGs were used for the next step.

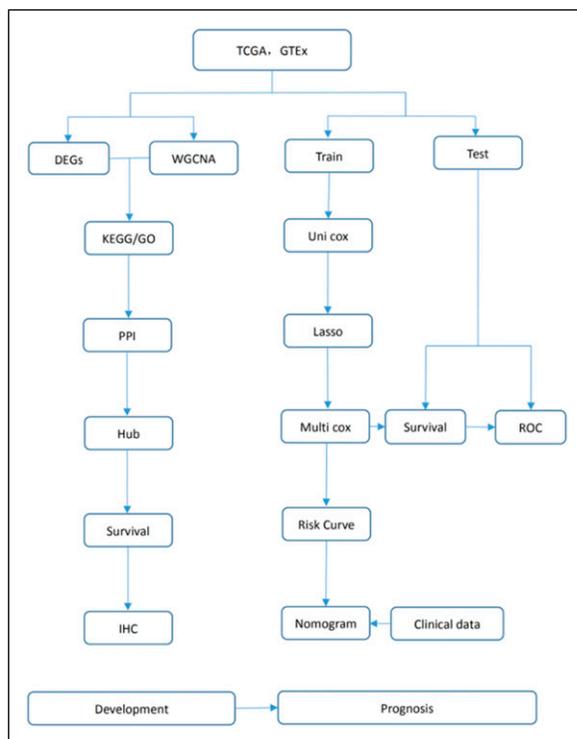


Figure 1. Experimental design.

## Kyoto Encyclopedia of Genes and Genomes, Gene Ontology, Protein-Protein Interaction Network, Survival Analyses, and Immunohistochemistry

The R package WGCNA was used to screen-out the relevant LUAD modules. The most relevant LUAD modules, genes intersecting with DEGs, and R package clusterprofiler were used for Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses of this part of the gene set. A PPI network for the intersection genes was constructed using the string database, and then the top 10 hub genes were screened using cytohubba in Cytoscape (version 3.7.2). The network PPI pairs with a combined confidence score of  $\geq 0.4$  were visualized. The GEPIA database (<http://gepia.cancer-pku.cn/>) survival analysis was performed for hub genes, which were screened for survival impact. The HPA database (<https://www.proteinatlas.org/>) was used to analyze the level of protein expression for hub genes.

### Prognostic Model Construction

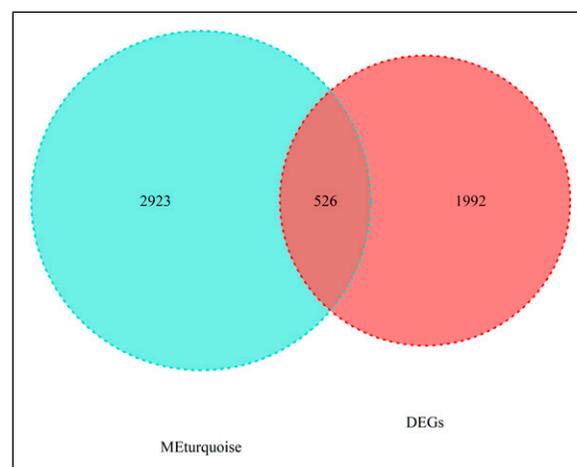
The mRNA results were obtained from the transcriptome data, the survival data were analyzed, the missing survival data were deleted, and 487 samples were obtained. The samples were randomly divided into 2 groups, including 243 training and 244 testing samples. R survival was used to perform a univariate Cox regression analysis on all genes in the training set to screen for genes significantly associated with prognosis.<sup>12</sup> To eliminate the problem of collinearity between genes, LASSO regression in R glmnet and R survival were used. After performing 1000 10-fold cross-validations, the  $\lambda$  value with minimized error was selected as the optimum  $\lambda$  parameter value.<sup>13</sup> Risk score models were constructed using multivariate Cox regression in R survminer and R survival. The risk score model's predicted risk score served as a predictor of prognostic status.

The risk score for each sample was calculated using a risk score model. The total sample was divided into high- and low-risk groups. Using R survminer and R survival were compared between high- and low-risk groups. To calculate the potency of the risk score model, ROC curves for the training and testing datasets were plotted using R survivalroc ROC. The risk curves were plotted according to survival time, survival status, gene expression quantity, and risk score for each sample. A nomogram was created using R RMS based on the clinical information provided by the TCGA database, including age, gender, stage, and risk score, and excluding missing data.

## Results

### Differentially Expressed Genes and Weighted Gene Co-expression Network Analysis

Based on the filter conditions of  $|\log_2 \text{FC}| \geq 2$  and adjusted  $P \leq .05$ , 2518 DEGs out of 19 405 genes were screened. The  $\beta$ -value was set at 12 (scale free  $R^2 \geq .90$ ) and adjusted as a



**Figure 2.** Intersection of 2518 differentially expressed genes and 3447 genes inside the MEturquoise module.

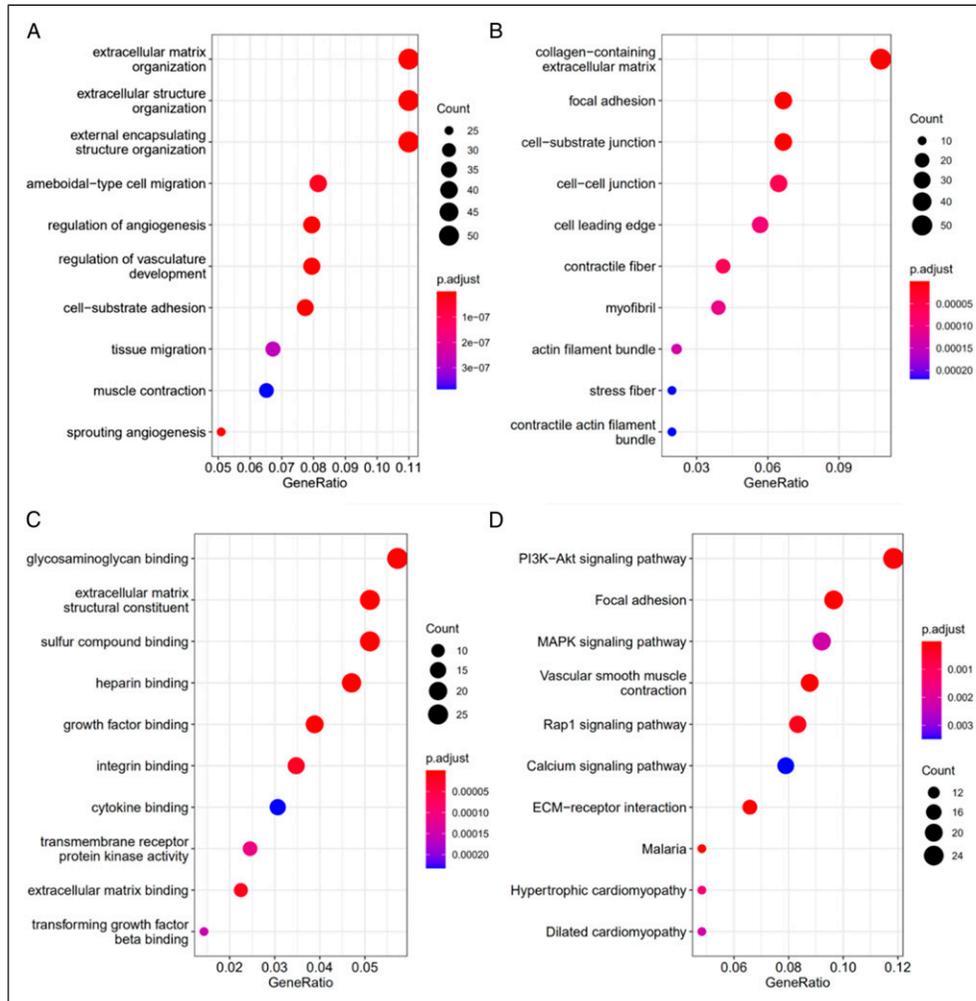
soft-threshold index to construct a scale-free co-expression network according to the WGCNA. The WGCNA results had a total of 10 modules, including MEred, MEblue, MEgreen, MEmagenta, MEpink, MEyellow, MEblack, MEBrown, MEturquoise, and MEGrey. MEturquoise was the module with the highest LUAD correlation with 3449 genes. It had a total of 526 intersection genes with DEGs (Figure 2).

### Enrichment Analysis

KEGG/GO functional enrichment analysis was performed on 526 genes. Biological process (BP) was mainly enriched in extracellular matrix organization, external structure organization, and external encapsulating structure organization (Figure 3A). Cellular component (CC) was mainly enriched in collagen-containing extracellular matrix, focal adhesion, and cell-substrate junction (Figure 3B). Molecular function (MF) was mainly enriched in glycosaminoglycan binding, extracellular matrix structural constituent, and sulfur compound binding (Figure 3C). KEGG was mainly enriched in the PI3K-Akt signaling pathway (Figure 3D).

### PPI, Hub Gene Survival Analysis, and Immunohistochemistry

A PPI network for 526 genes was constructed using a string database. The top 10 hub genes were screened using the cytohubba module in Cytoscape. These included IL6, PECAM1, CDH5, VWF, THBS1, CAV1, TEK, HGF, SPP1, and *ENG* (Figure 4A). Survival analysis for hub genes identified 4 genes that had an impact on survival. They were PECAM1, HGF, SPP1, and *ENG* (Figure 4B). Boxplots based on the expression profiles of these 4 genes were constructed (Figure 4C). PECAM1, HGF, and *ENG* were downregulated and SPP1 was upregulated at the



**Figure 3.** A. Top 10 GO terms of genes related to biological process, B. cellular component, C. molecular function, D. kyoto encyclopedia of genes and genomes.

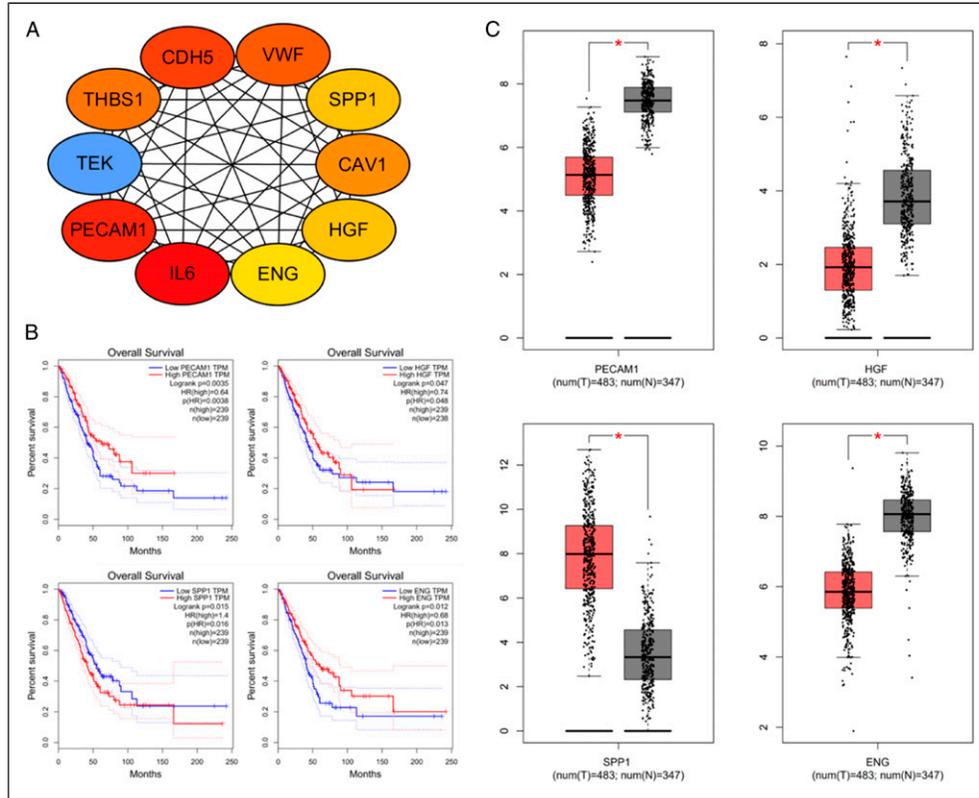
mRNA level. Immunohistochemistry results for PE-CAM1, HGF, and *ENG* showed that they were down-regulated and *SPP1* was upregulated at the protein level (Figure 5).

### Prognostic Model Construction

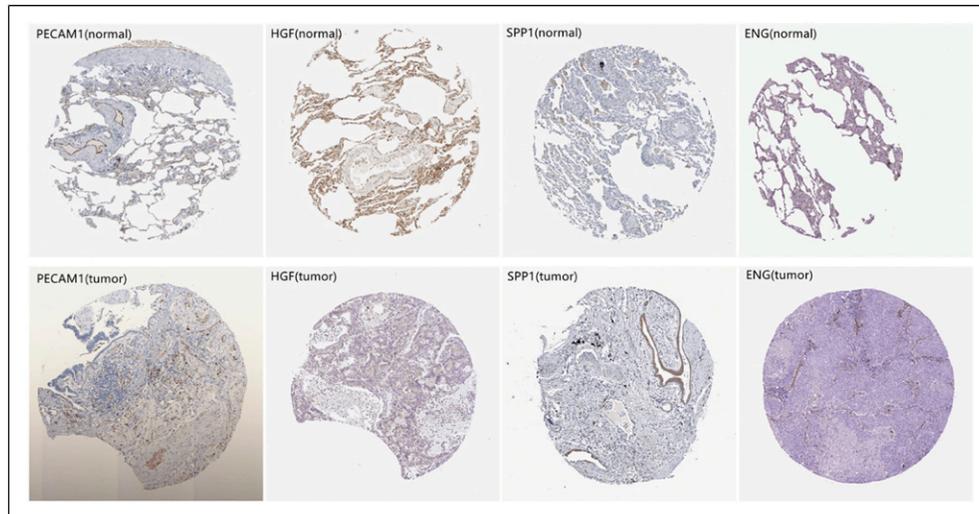
An accurate prognostic model was established and univariate Cox regression analysis was performed on 19 405 genes in 243 training dataset samples, which identified 75 prognostically influential genes. LASSO regression analysis was performed on 75 genes, which identified 17 genes. Multivariate Cox regression analysis of 17 genes filtered out 9 genes (*RPL22*, *ENO1*, *KDF1*, *ZNF691*, *PCSK9*, *DNASE2B*, *SNX7*, *ELAPOR1*, and *LCE5A*) that were used to construct a forest plot (Figure 6A). The favorable prognosis genes included *KDF1*, *ZNF691*, *DNASE2B*, and *ELAPOR1*, while unfavorable prognosis genes included *RPL22*, *ENO1*, *PCSK9*, *SNX7*, and *LCE5A*. The risk score model followed the relationship,

where  $\text{riskScore} = .421 \cdot \text{RPL22} + .417 \cdot \text{ENO1} - .522 \cdot \text{KDF1} - .502 \cdot \text{ZNF691} + .169 \cdot \text{PCSK9} - .13 \cdot \text{DNASE2B} + .346 \cdot \text{SNX7} - .077 \cdot \text{ELAPOR1} + .179 \cdot \text{LCE5A}$ . Risk score analysis for the training and testing datasets divided the cohort into high- and low-risk groups according to the median risk score value. Survival analysis showed that the high risk group had a worse overall survival than the low-risk group ( $P < .0001$ ; Figure 6B-6C).

The ROC curves for the training and testing datasets were plotted to evaluate the efficacy of the risk score model. The ROC curves had area under the curve values of .78 and .758 for the training and testing groups, respectively (Figure 6D-6E). Risk curves were plotted according to survival time and risk score (Figure 7A). The number of patients who died continuously increased and the number of those who survived continuously decreased with increasing risk scores (Figure 7B). A heat map of the key genes for prognosis is shown in Figure 7C, which demonstrated the reliability of the risk score



**Figure 4.** A. Top 10 hub genes were screened from 526 genes. B. Top 10 genes were subjected to survival analysis, and 4 genes (PECAM1, HGF, SPP1, and ENG) had impacts on survival with a log-rank of  $P < .05$ . C. Expression profiles of PECAM1, HGF, SPP1, and ENG in LUAD; SPP1 was upregulated, while PECAM1, HGF, and ENG were downregulated. \*Indicates that the difference between LUAD and normal groups was statistically significant.

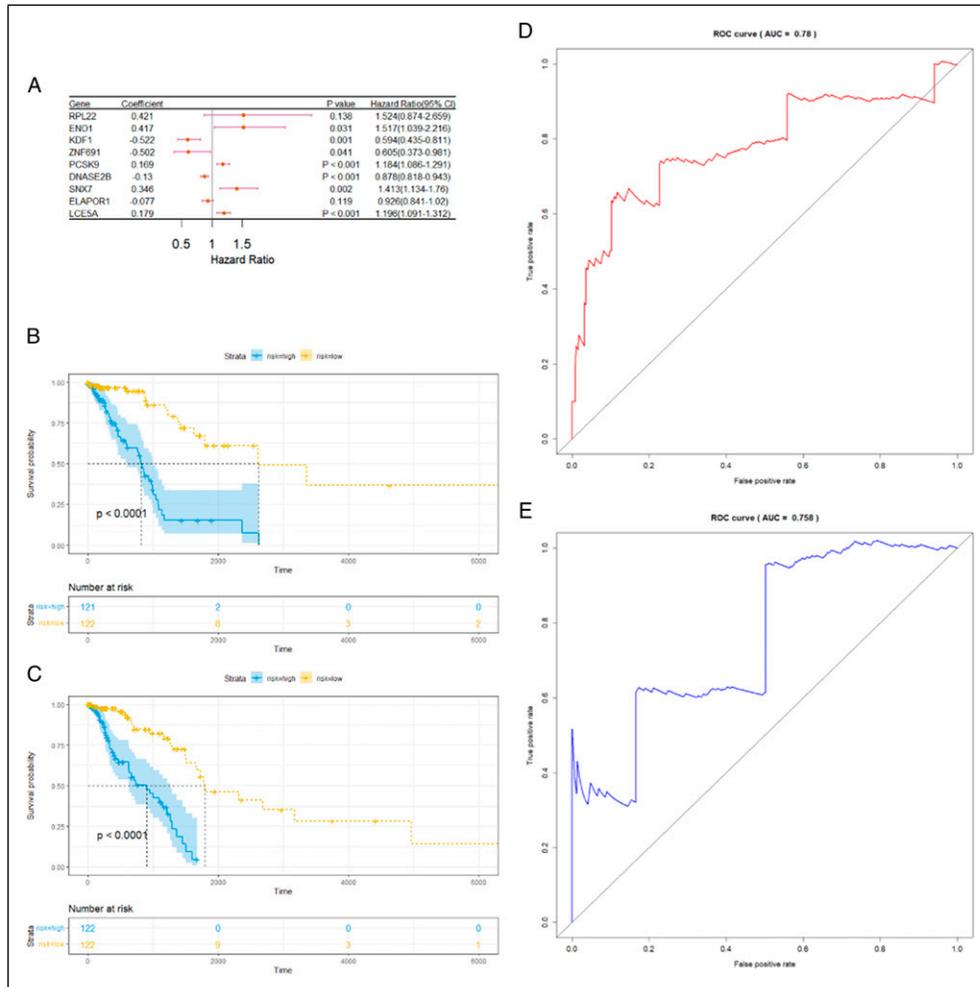


**Figure 5.** PECAM1, HGF, and ENG were downregulated and SPP1 was upregulated at the protein level.

model. The nomogram with clinical data from the TCGA database included age, gender, stage, and risk score and was used to predict the 1-, 3-, and 5-year survival values (Figure 8).

## Discussion

The present study used DEGs with WGCNA to identify the intersection genes. KEGG/GO enrichment analysis was

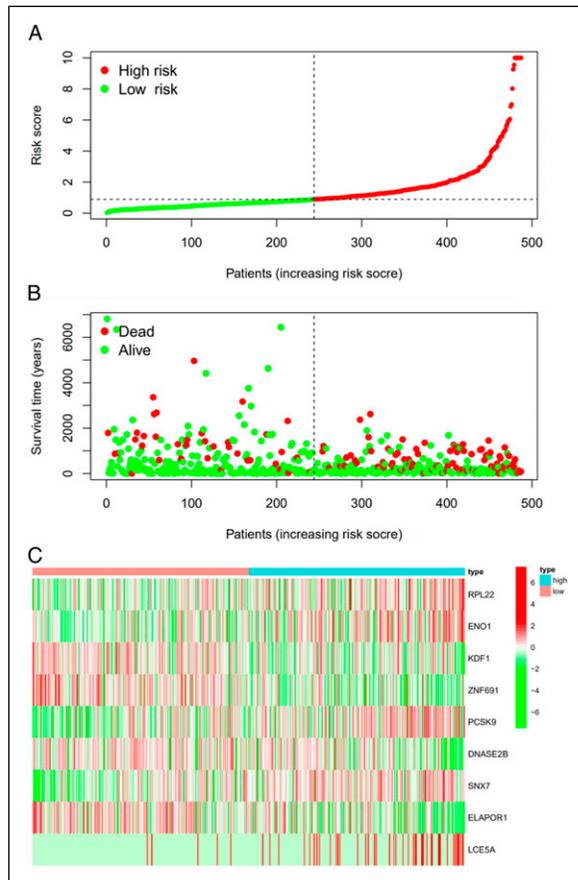


**Figure 6.** A. KDF1, ZNF691, DNASE2B, and ELAPOR1 had coefficient values of  $<1$  and were favorable prognostic genes. RPL22, ENO1, PCSK9, SNX7, and LCE5A had coefficient values of  $>1$  and were unfavorable prognostic genes. B. The high-risk group had lower overall survival than the low-risk group in the training dataset ( $P < .0001$ ). C. The high-risk group had a lower overall survival rate than the low-risk group in the testing dataset ( $P < .0001$ ). Risk model potency calculated by plotting ROC curves with area under the curve values of .78 and .758 for training D and testing E groups, respectively.

performed on the intersection genes, which were then used to construct the PPI network. Ten hub genes were screened out and evaluated using the cytohubba module in Cytoscape. The final results screened-out 4 genes that have effects on the development of LUAD. KEGG analysis showed that PECAM-1 participates in cell adhesion molecules, leukocyte trans-endothelial migration, and malaria. Increased expression of PECAM-1 promotes migration of endothelial lymphocytes.<sup>14</sup> The present study demonstrated that high expression of PECAM-1 was associated with improved survival, which was consistent with the study by Cao et al<sup>15</sup> PECAM-1 signaling may be involved in the regulation of VEGF expression<sup>16</sup> HGF and SPP1 were involved in the PI3K-Akt signaling pathway. The present study demonstrated that high expression of HGF was associated with improved survival, which was consistent with the study by Pan et al<sup>17</sup> It has also been noted that HGF regulates the expression of SAE2 and circRNA CCDC66,

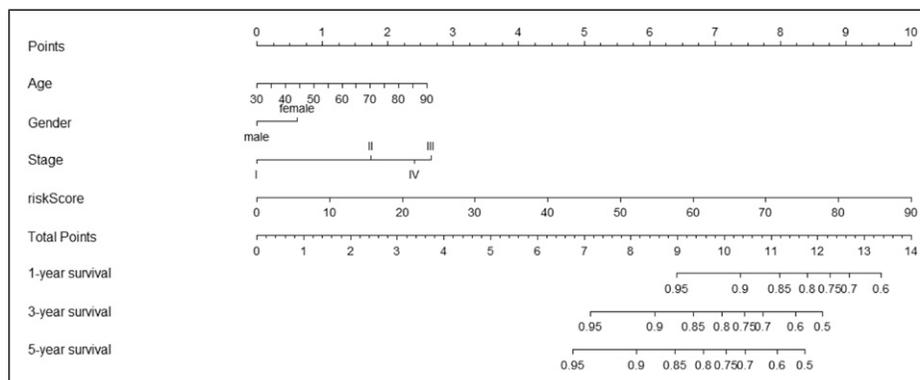
thereby increasing EMT and drug resistance in LADC cells.<sup>18</sup> Additional studies have demonstrated that HGF activates FAK and downregulates the expression of AIF.<sup>19</sup> The cDNA sequence for SPP1 contains a 67-bp 5' noncoding region and a 415-bp 3' noncoding region, as well as a 942-bp coding region encoding a 314-amino-acid protein.<sup>20</sup> High SPP1 expression predicts poor survival in LUAD,<sup>21</sup> and SPP1 serves as a biological marker for LUAD prognosis.<sup>22-25</sup> It is also involved in the development of pulmonary fibrosis, as well as affects macrophage secretion by upregulating its expression.<sup>26</sup> The membrane antigen endoglin, which was first identified in 1985 on the surface of a cell line with acute lymphoblastic leukaemia using hybrid technology to obtain the murine monoclonal antibody mab44g4, was mainly localized in vascular endothelial cells.<sup>27</sup> In 1988, Gougos et al<sup>28</sup> have validated the presence of endoglin in a variety of vascular endothelial cells using the monoclonal antibody mab44g4 and have found it to

be a homodimer composed of subunits with a molecular weight of 95 kDa. Mutations in *ENG* may cause hereditary hemorrhagic telangiectasia,<sup>29,30</sup> greater susceptibility to pulmonary artery malformations,<sup>31</sup> and pulmonary hypertension.<sup>32</sup>



**Figure 7.** A. High- and low-risk groups were ranked by risk score of each sample. B. As risk score increased, the number of deaths continued to increase in each sample. C. Expression of each gene inside the risk score model in high- and low-risk groups.

Univariate Cox, LASSO, and multivariate Cox regression analyses were used in the present study to construct a risk score model. The model contained 9 genes, including RPL22, ENO1, KDF1, ZNF691, *PCSK9*, DNASE2B, SNX7, ELAPOR1, and LCE5A. Ribosomal protein L22 (rpl22) is expressed in various cells.<sup>33</sup> Studies have shown that rpl22 functions as a tumor suppressor by selectively upregulating the expression of the tumor suppressor p53 and inhibiting colony formation of cancer cells.<sup>34</sup> The rpl22 is down-regulated in lung cancer and can interact with casein kinase 2 $\alpha$ .<sup>35,36</sup> ENO1 is the most differentially expressed protein in humans, and this differential expression depends on the stress or metabolic status of the pathological cells in the tissue.<sup>37</sup> Additionally, ENO1 is involved in many important physiological processes, such as hypoxia resistance,<sup>38</sup> inflammatory responses, and autoimmune activities.<sup>39</sup> ENO1 upregulation promotes glycolysis and tumor progression in LUAD<sup>40</sup> and serves as a potential biological marker in the early occurrence and development of lung cancer.<sup>41</sup> Chemoresistance influences small cell lung cancer by regulating ENO1 expression.<sup>42</sup> The KDF1 mutation may cause abnormal ectodermal development<sup>43</sup> and regulates epidermal differentiation via KDF1-mediated IKK $\alpha$  deubiquitination.<sup>44</sup> The low expression of ZNF691 in patients with ovarian cancer is associated with poor prognosis.<sup>45</sup> Proprotein convertase subtilisin 9 (*PCSK9*) is the ninth member of the proteinase K subfamily of the proprotein convertase family. It is currently confirmed that *PCSK9* can affect blood lipids by regulating lipid metabolism. It interacts primarily at the cell surface with epidermal growth factor precursor homology domain A of the *LDLR*, limiting the LDLR-mediated lipoprotein uptake.<sup>46</sup> *PCSK9* regulates apoptosis of LUAD cells (A549) via the ER stress and mitochondrial signaling pathways.<sup>47</sup> It may also be a prognostic marker for NSCLC patients.<sup>48</sup> DNASE2B is an enzyme responsible for nuclear degradation in the mouse lens. However, DNASE2B expression in zebrafish has a distribution pattern that is



**Figure 8.** “Points” served as a scoring scale for each factor, while “total points” served as a scale for the total score. “riskScore” represents the risk score. “Stage” represents the 4 stages of LUAD. Based on the total score of each patient, 1-, 3-, and 5-year survival rates were inferred.

different from that in mice.<sup>49</sup> DNASE2B is highly expressed in smokers and may have the potential to improve the prediction of chemosensitivity in gastric cancer patients.<sup>50,51</sup> Sorting nexins (SNXs) are a family of peripheral membrane proteins that direct protein trafficking decisions within the endocytic network.<sup>52</sup> SNX7 is a member of the SNX family and contains a PX and BAR domains. It regulates amyloid  $\beta$  peptide by inducing the degradation of amyloid precursor protein.<sup>53</sup> ELAPOR1 is a secretory granule maturation-promoting factor that is lost during paligenosis.<sup>54</sup>

The study results showed that as the risk score increased, the patient survival rate gradually decreased, while the prognostic impact of the risk score increased. The area under the ROC curve for the training and testing datasets indicated that the risk score model was valid.

## Limitations

The present study did not validate the amount of mRNA expression of PECAM1, HGF, SPPI, and *ENG* in lung cancer and only adopted the mRNA data from the TCGA database to construct a prognostic model with a single source of data.

## Conclusions

Genes affecting the development of LUAD included PECAM1, HGF, SPP1, and *ENG*. The favorable prognosis genes included KDF1, ZNF691, DNASE2B, and ELAPOR1, while the unfavorable prognosis genes included RPL22, ENO1, *PCSK9*, SNX7, and LCE5A.

## Acknowledgments

Thanks TCGA, GETx for providing raw data.

## Author Contribution

Xuan Luo and Li Bian designed the research study. Jian Guo Xu analyzed the data. ZhiYuan Wang, XiaoFang Wang, QianYing Zhu and Juan Zhao wrote the paper.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Regional Fund Project of the National Natural Science Foundation of China and Yunnan Provincial Department of Science and Technology-Kunming Medical University applied basic research joint special major project (Serial number: 82060423; 202001AY0700 1-007).

## Ethical Approval

Ethical approval was not necessary for this study because data were obtained from public databases.

## ORCID iDs

Luo Xuan  <https://orcid.org/0000-0002-7838-2489>

Bian Li  <https://orcid.org/0000-0001-9916-3720>

Xu Jian Guo  <https://orcid.org/0000-0002-6118-2413>

Wang ZhiYuan  <https://orcid.org/0000-0002-9115-8676>

Wang XiaoFang  <https://orcid.org/0000-0003-3275-5865>

Zhu QianYing  <https://orcid.org/0000-0002-7463-5290>

Zhao Juan  <https://orcid.org/0000-0002-3132-2104>

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin*. 2018;68(6):394-424.
2. Hirsch FR, Scagliotti GV, et al. Lung cancer: Current therapies and new targeted treatments. *Lancet*. 2017;389(10066):299-311.
3. Goodgame B, Viswanathan A, Miller CR, et al. A clinical model to estimate recurrence risk in resected stage I non-small cell lung cancer. *Am J Clin Oncol*. 2008;31(1):22-28.
4. Wu T-H, Hsiue EH-C, Lee J-H, et al. Best response according to RECIST during first-line EGFR-TKI treatment predicts survival in EGFR mutation-positive non-small-cell lung cancer patients. *Clin Lung Cancer*. 2018;19(3):e361-e372.
5. Zhang M, Wang Q, Ding Y, et al. CUX1-ALK, a Novel ALK rearrangement that responds to Crizotinib in non-small cell lung cancer. *J Thorac Oncol*. 2018;13(11):1792-1797.
6. Gao M, Kong W, Huang Z, Xie Z. Identification of key genes related to lung squamous cell carcinoma using bioinformatics analysis. *Int J Mol Sci*. 2020;21(8):2994.
7. Meng Z, Xiaojun W, Hongbo S, et al. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer [J]. *Oncotarget*. 2016;7(11):12598-12611.
8. Udyavar AR, Hoeksema MD, Clark JE, et al. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol*. 2013;7(Suppl 5):S1.
9. Tang J, Kong D, Cui Q, et al. Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front Oncol*. 2018;8(8):374-374.
10. Wei Z, Zhongqiu T, Lu S, Zhang F, Xie W, Wang Y. Gene coexpression analysis offers important modules and pathway of human lung adenocarcinomas. *J Cell Physiol*. 2020;235(1):454-464.
11. Yi M, Li T, Qin S, et al. Identifying tumorigenesis and prognosis-related genes of lung adenocarcinoma: Based on weighted gene coexpression network analysis. *BioMed Res Int*. 2020;2020:1-15.
12. Lunn M, McNeil D. Applying cox regression to competing risks. *Biometrics*. 1995;51(2):524-532.

13. Shahraki HR, Salehi A, Zare N. Survival prognostic factors of male breast cancer in Southern Iran: A LASSO-Cox regression approach. *Asian Pac J Cancer Prev APJCP*. 2015;16(15):6773-6777.
14. Dziejulska D, Drac H, Micej W, Mieszkowski J, Rafałowska J. Paraneoplastic syndrome in the course of lung adenocarcinoma: morphological picture and immunohistochemical analysis of the inflammatory infiltrates and PECAM-1 expression. *Folia Neuropathol*. 2000;38(1):29-33.
15. Cao S, Wang Y, Li J, et al. Prognostic implication of the expression level of PECAM-1 in non-small cell lung cancer. *Front Oncol*. 2021;11:587744.
16. Ouyang J-S, Li Y-P, Chen C-S, et al. Inhibition of lung tumor growth in nude mice by siRNACD31 targeting PECAM-1. *Oncol Lett*. 2014;8(1):33-40.
17. Pan B, Wang R, Zhang J, Chen H, Huang Y, Garfield D. HGF and NRG1 protein expression are not poor prognostic markers in surgically resected lung adenocarcinoma. *OncoTargets Ther*. 2015;8:1185-1191.
18. Joseph NA, Chiou S-H, Lung Z, et al. The role of HGF-MET pathway and CCDC66 cirRNA expression in EGFR resistance and epithelial-to-mesenchymal transition of lung adenocarcinoma cells. *J Hematol Oncol*. 2018;11(1):74.
19. Chen J-T, Huang C-Y, Chiang Y-Y, et al. HGF increases cisplatin resistance via down-regulation of AIF in lung cancer cells. *Am J Respir Cell Mol Biol*. 2008;38(5):559-565.
20. Kiefer MC, Bauer DM, Barr PJ. The cDNA and derived amino acid sequence for human osteopontin. *Nucleic Acids Res*. 1989;17(8):3306-3306.
21. Giopanou I, Kanellakis NI, Giannou AD, et al. Osteopontin drives KRAS-mutant lung adenocarcinoma. *Carcinogenesis*. 2020;41(8):1134-1144.
22. Tu Y, Chen C, Fan G. Association between the expression of secreted phosphoprotein - related genes and prognosis of human cancer. *BMC Cancer*. 2019;19(1):1230.
23. Guo Z, Huang J, Wang Y, et al. Analysis of expression and its clinical significance of the secreted phosphoprotein 1 in lung adenocarcinoma. *Front Genet*. 2020;11:547.
24. Su C, Liu W-X, Wu L-S, Dong T-J, Liu J-F. Screening of Hub gene targets for lung cancer via microarray data. *Comb Chem High Throughput Screen*. 2021;24(2):269-285.
25. Shen XY, Liu XP, Song CK, Wang YJ, Li S, Hu WD. Genome-wide analysis reveals alcohol dehydrogenase 1C and secreted phosphoprotein 1 for prognostic biomarkers in lung adenocarcinoma. *J Cell Physiol*. 2019;234(12):22311-22320.
26. Wang H, Wang M, Xiao K, et al. Bioinformatics analysis on differentially expressed genes of alveolar macrophage in IPF. *Exp Lung Res*. 2019;45(9-10):288-296.
27. Quackenbush EJ, Letarte M. Identification of several cell surface proteins of non-T, non-B acute lymphoblastic leukemia by using monoclonal antibodies. *J Immunol*. 1985;134(2):1276-1285. PMID: 3155538.
28. Gougos A, Letarte M. Identification of a human endothelial cell antigen with monoclonal antibody 44G4 produced against a pre-B leukemic cell line. *J Immunol*. 1988;141(6):1925-1933. PMID: 3262644.
29. Villa D, Cinnante C, Valcamonica G, et al. Hereditary hemorrhagic telangiectasia associated with cortical development malformation due to a start loss mutation in ENG. *BMC Neurol*. 2020;20(1):316.
30. Damjanovich K, Langa C, Blanco FJ, et al. 5'UTR mutations of ENG cause hereditary hemorrhagic telangiectasia. *Orphanet J Rare Dis*. 2011;6:85.
31. Mu W, Cordner ZA, Yuqi Wang K, et al. Characterization of pulmonary arteriovenous malformations in ACVRL1 versus ENG mutation carriers in hereditary hemorrhagic telangiectasia. *Genet Med*. 2018;20(6):639-644.
32. Pousada G, Baloiira A, Fontán D, Núñez M, Valverde D. Mutational and clinical analysis of the ENG gene in patients with pulmonary arterial hypertension. *BMC Genet*. 2016;17(1):72.
33. Solanki NR, Stadanlick JE, Zhang Y, Duc A-C, Lee S-Y, Lauritsen JPH, et al. Rpl22 loss selectively impairs  $\alpha\beta$  T cell development by dysregulating endoplasmic reticulum stress signaling. *J Immunol*. 2016;197(6):2280-2289.
34. Cao B, Fang Z, Liao P, et al. Cancer-mutated ribosome protein L22 (RPL22/eL22) suppresses cancer cell survival by blocking p53-MDM2 circuit. *Oncotarget*. 2017;8(53):90651-90661.
35. Yang M, Sun H, He J, et al. Interaction of ribosomal protein L22 with casein kinase 2 $\alpha$ : A novel mechanism for understanding the biology of non-small cell lung cancer. *Oncol Rep*. 2014;32(1):139-144.
36. Yang M, Sun H, Wang H, Zhang S, Yu X, Zhang L. Down-regulation of ribosomal protein L22 in non-small cell lung cancer. *Med Oncol*. 2013;30(3):646.
37. Petrak J, Ivanek R, Toman O, et al. Déjà vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. *Proteomics*. 2008;8(9):1744-1749.
38. Lung J, Liu K-J, Chang J-Y, Leu S-J, Shih N-Y. MBP-1 is efficiently encoded by an alternative transcript of the ENO1 gene but post-translationally regulated by proteasome-dependent protein turnover. *FEBS J*. 2010;277(20):4308-4321.
39. Amedei A, Niccolai E, Benagiano M, et al. Ex vivo analysis of pancreatic cancer-infiltrating T lymphocytes reveals that ENO-specific Tregs accumulate in tumor tissue and inhibit Th1/Th17 effector cell functions. *Cancer Immunol Immunother*. 2013;62(7):1249-1260.
40. Zhou J, Zhang S, Chen Z, He Z, Xu Y, Li Z. CircRNA-ENO1 promoted glycolysis and tumor progression in lung adenocarcinoma through upregulating its host gene ENO1. *Cell Death Dis*. 2019;10(12):885.
41. Yu L, Shen J, Mannoor K, Guarnera M, Jiang F. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics. *Clin Lung Cancer*. 2014;15(5):372-378.
42. Chen R, Li D, Zheng M, et al. FGFRL1 affects chemoresistance of small-cell lung cancer by modulating the PI3K/Akt pathway via ENO1. *J Cell Mol Med*. 2020;24(3):2123-2134.

43. Manaspon C, Thaweesapphithak S, Osathanon T, Suphapeetiporn K, Porntaveetus T, Shotelersuk V. A novel de novo mutation substantiates KDF 1 as a gene causing ectodermal dysplasia. *Br J Dermatol*. 2019;181(2):419-420.
44. Li Y, Tang L, Yue J, et al. Regulation of epidermal differentiation through KDF1-mediated deubiquitination of IKK $\alpha$ . *EMBO Reports*. 2020;21(5):e48566.
45. Yamanoi K, Baba T, Abiko K, et al. Acquisition of a side population fraction augments malignant phenotype in ovarian cancer. *Sci Rep*. 2019;9(1):14215.
46. Gu H-m, Adijiang A, Mah M, Zhang D-w. Characterization of the role of EGF-A of low density lipoprotein receptor in PCSK9 binding. *JLR (J Lipid Res)*. 2013;54(12):3345-3357.
47. Xu X, Cui Y, Cao L, Zhang Y, Yin Y, Hu X. PCSK9 regulates apoptosis in human lung adenocarcinoma A549 cells via endoplasmic reticulum stress and mitochondrial signaling pathways. *Exp Ther Med*. 2017;13(5):1993-1999.
48. Bonaventura A, Grossi F, Montecucco F. PCSK9 is a promising prognostic marker in patients with advanced NSCLC. *Cancer Immunol Immunother*. 2020;69(3):491-492.
49. Iida A, Tabata Y, Baba Y, Fujii T, Watanabe S. Critical roles of DNase1131 in lens nuclear degeneration in zebrafish. *Biochimie*. 2014;106:68-74.
50. Ha YJ, Yoon SN, Jeon YJ, et al. Genome-wide identification of chemosensitive single nucleotide polymorphism markers in gastric cancer. *Anticancer Res*. 2011;31(12):4329-4338.
51. Pintarelli G, Noci S, Maspero D, et al. Cigarette smoke alters the transcriptome of non-involved lung tissue in lung adenocarcinoma patients. *Sci Rep*. 2019;9(1):13039.
52. Antón Z, Betin VMS, Simonetti B, et al. A heterodimeric SNX4–SNX7 SNX-BAR autophagy complex coordinates ATG9A trafficking for efficient autophagosome assembly. *J Cell Sci*. 2020;133(14):jcs246306.
53. Xu S, Zhang L, Brodin L. Overexpression of SNX7 reduces A $\beta$  production by enhancing lysosomal degradation of APP. *Biochem Biophys Res Commun*. 2018;495(1):12-19.
54. Cho CJ, Park D, Mills JC. ELAPOR1 is a secretory granule maturation-promoting factor that is lost during paligenesis. *Am J Physiol Gastrointest Liver Physiol*. 2022;322(1):G49-G65.