

World Journal of *Clinical Cases*

World J Clin Cases 2022 June 26; 10(18): 5934-6340



MINIREVIEWS

- 5934 Development of clustered regularly interspaced short palindromic repeats/CRISPR-associated technology for potential clinical applications
Huang YY, Zhang XY, Zhu P, Ji L
- 5946 Strategies and challenges in treatment of varicose veins and venous insufficiency
Gao RD, Qian SY, Wang HH, Liu YS, Ren SY
- 5957 Diabetes mellitus susceptibility with varied diseased phenotypes and its comparison with phenome interactome networks
Rout M, Kour B, Vuree S, Lulu SS, Medicherla KM, Suravajhala P

ORIGINAL ARTICLE**Clinical and Translational Research**

- 5965 Identification of potential key molecules and signaling pathways for psoriasis based on weighted gene co-expression network analysis
Shu X, Chen XX, Kang XD, Ran M, Wang YL, Zhao ZK, Li CX
- 5984 Construction and validation of a novel prediction system for detection of overall survival in lung cancer patients
Zhong C, Liang Y, Wang Q, Tan HW, Liang Y

Case Control Study

- 6001 Effectiveness and postoperative rehabilitation of one-stage combined anterior-posterior surgery for severe thoracolumbar fractures with spinal cord injury
Zhang B, Wang JC, Jiang YZ, Song QP, An Y

Retrospective Study

- 6009 Prostate sclerosing adenopathy: A clinicopathological and immunohistochemical study of twelve patients
Feng RL, Tao YP, Tan ZY, Fu S, Wang HF
- 6021 Value of magnetic resonance diffusion combined with perfusion imaging techniques for diagnosing potentially malignant breast lesions
Zhang H, Zhang XY, Wang Y
- 6032 Scar-centered dilation in the treatment of large keloids
Wu M, Gu JY, Duan R, Wei BX, Xie F
- 6039 Application of a novel computer-assisted surgery system in percutaneous nephrolithotomy: A controlled study
Qin F, Sun YF, Wang XN, Li B, Zhang ZL, Zhang MX, Xie F, Liu SH, Wang ZJ, Cao YC, Jiao W

- 6050** Influences of etiology and endoscopic appearance on the long-term outcomes of gastric antral vascular ectasia

Kwon HJ, Lee SH, Cho JH

Randomized Controlled Trial

- 6060** Evaluation of the clinical efficacy and safety of TST33 mega hemorrhoidectomy for severe prolapsed hemorrhoids

Tao L, Wei J, Ding XF, Ji LJ

- 6069** Sequential chemotherapy and icotinib as first-line treatment for advanced epidermal growth factor receptor-mutated non-small cell lung cancer

Sun SJ, Han JD, Liu W, Wu ZY, Zhao X, Yan X, Jiao SC, Fang J

Randomized Clinical Trial

- 6082** Impact of preoperative carbohydrate loading on gastric volume in patients with type 2 diabetes

Lin XQ, Chen YR, Chen X, Cai YP, Lin JX, Xu DM, Zheng XC

META-ANALYSIS

- 6091** Efficacy and safety of adalimumab in comparison to infliximab for Crohn's disease: A systematic review and meta-analysis

Yang HH, Huang Y, Zhou XC, Wang RN

CASE REPORT

- 6105** Successful treatment of acute relapse of chronic eosinophilic pneumonia with benralizumab and without corticosteroids: A case report

Izhakian S, Pertzov B, Rosengarten D, Kramer MR

- 6110** Pembrolizumab-induced Stevens-Johnson syndrome in advanced squamous cell carcinoma of the lung: A case report and review of literature

Wu JY, Kang K, Yi J, Yang B

- 6119** Hepatic epithelioid hemangioendothelioma after thirteen years' follow-up: A case report and review of literature

Mo WF, Tong YL

- 6128** Effectiveness and safety of ultrasound-guided intramuscular lauromacrogol injection combined with hysteroscopy in cervical pregnancy treatment: A case report

Ye JP, Gao Y, Lu LW, Ye YJ

- 6136** Carcinoma located in a right-sided sigmoid colon: A case report

Lyu LJ, Yao WW

- 6141** Subcutaneous infection caused by *Mycobacterium abscessus* following cosmetic injections of botulinum toxin: A case report

Deng L, Luo YZ, Liu F, Yu XH

- 6148** Overlapping syndrome of recurrent anti-N-methyl-D-aspartate receptor encephalitis and anti-myelin oligodendrocyte glycoprotein demyelinating diseases: A case report
Yin XJ, Zhang LF, Bao LH, Feng ZC, Chen JH, Li BX, Zhang J
- 6156** Liver transplantation for late-onset ornithine transcarbamylase deficiency: A case report
Fu XH, Hu YH, Liao JX, Chen L, Hu ZQ, Wen JL, Chen SL
- 6163** Disseminated strongyloidiasis in a patient with rheumatoid arthritis: A case report
Zheng JH, Xue LY
- 6168** CYP27A1 mutation in a case of cerebrotendinous xanthomatosis: A case report
Li ZR, Zhou YL, Jin Q, Xie YY, Meng HM
- 6175** Postoperative multiple metastasis of clear cell sarcoma-like tumor of the gastrointestinal tract in adolescent: A case report
Huang WP, Li LM, Gao JB
- 6184** Toripalimab combined with targeted therapy and chemotherapy achieves pathologic complete response in gastric carcinoma: A case report
Liu R, Wang X, Ji Z, Deng T, Li HL, Zhang YH, Yang YC, Ge SH, Zhang L, Bai M, Ning T, Ba Y
- 6192** Presentation of Boerhaave's syndrome as an upper-esophageal perforation associated with a right-sided pleural effusion: A case report
Tan N, Luo YH, Li GC, Chen YL, Tan W, Xiang YH, Ge L, Yao D, Zhang MH
- 6198** Camrelizumab-induced anaphylactic shock in an esophageal squamous cell carcinoma patient: A case report and review of literature
Liu K, Bao JF, Wang T, Yang H, Xu BP
- 6205** Nontraumatic convexal subarachnoid hemorrhage: A case report
Chen HL, Li B, Chen C, Fan XX, Ma WB
- 6211** Growth hormone ameliorates hepatopulmonary syndrome and nonalcoholic steatohepatitis secondary to hypopituitarism in a child: A case report
Zhang XY, Yuan K, Fang YL, Wang CL
- 6218** Vancomycin dosing in an obese patient with acute renal failure: A case report and review of literature
Xu KY, Li D, Hu ZJ, Zhao CC, Bai J, Du WL
- 6227** Insulinoma after sleeve gastrectomy: A case report
Lobaton-Ginsberg M, Sotelo-González P, Ramirez-Renteria C, Juárez-Aguilar FG, Ferreira-Hermosillo A
- 6234** Primary intestinal lymphangiectasia presenting as limb convulsions: A case report
Cao Y, Feng XH, Ni HX
- 6241** Esophagogastric junctional neuroendocrine tumor with adenocarcinoma: A case report
Kong ZZ, Zhang L

- 6247** Foreign body granuloma in the tongue differentiated from tongue cancer: A case report
Jiang ZH, Xu R, Xia L
- 6254** Modified endoscopic ultrasound-guided selective N-butyl-2-cyanoacrylate injections for gastric variceal hemorrhage in left-sided portal hypertension: A case report
Yang J, Zeng Y, Zhang JW
- 6261** Management of type IIIb dens invaginatus using a combination of root canal treatment, intentional replantation, and surgical therapy: A case report
Zhang J, Li N, Li WL, Zheng XY, Li S
- 6269** Clivus-involved immunoglobulin G4 related hypertrophic pachymeningitis mimicking meningioma: A case report
Yu Y, Lv L, Yin SL, Chen C, Jiang S, Zhou PZ
- 6277** De novo brain arteriovenous malformation formation and development: A case report
Huang H, Wang X, Guo AN, Li W, Duan RH, Fang JH, Yin B, Li DD
- 6283** Coinfection of *Streptococcus suis* and *Nocardia asiatica* in the human central nervous system: A case report
Chen YY, Xue XH
- 6289** Dilated left ventricle with multiple outpouchings – a severe congenital ventricular diverticulum or left-dominant arrhythmogenic cardiomyopathy: A case report
Zhang X, Ye RY, Chen XP
- 6298** Spontaneous healing of complicated crown-root fractures in children: Two case reports
Zhou ZL, Gao L, Sun SK, Li HS, Zhang CD, Kou WW, Xu Z, Wu LA
- 6307** Thyroid follicular renal cell carcinoma excluding thyroid metastases: A case report
Wu SC, Li XY, Liao BJ, Xie K, Chen WM
- 6314** Appendiceal bleeding: A case report
Zhou SY, Guo MD, Ye XH
- 6319** Spontaneous healing after conservative treatment of isolated grade IV pancreatic duct disruption caused by trauma: A case report
Mei MZ, Ren YF, Mou YP, Wang YY, Jin WW, Lu C, Zhu QC
- 6325** Pneumonia and seizures due to hypereosinophilic syndrome – organ damage and eosinophilia without synchronisation: A case report
Ishida T, Murayama T, Kobayashi S
- 6333** Creutzfeldt-Jakob disease presenting with bilateral hearing loss: A case report
Na S, Lee SA, Lee JD, Lee ES, Lee TK

LETTER TO THE EDITOR

- 6338** Stem cells as an option for the treatment of COVID-19
Cuevas-González MV, Cuevas-González JC

ABOUT COVER

Editorial Board Member of *World Journal of Clinical Cases*, Cristina Tudoran, PhD, Assistant Professor, Department VII, Internal Medicine II, Discipline of Cardiology, "Victor Babes" University of Medicine and Pharmacy Timisoara, Timisoara 300041, Timis, Romania. cristina13.tudoran@gmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Clinical Cases* (*WJCC*, *World J Clin Cases*) is to provide scholars and readers from various fields of clinical medicine with a platform to publish high-quality clinical research articles and communicate their research findings online.

WJCC mainly publishes articles reporting research results and findings obtained in the field of clinical medicine and covering a wide range of topics, including case control studies, retrospective cohort studies, retrospective studies, clinical trials studies, observational studies, prospective studies, randomized controlled trials, randomized clinical trials, systematic reviews, meta-analysis, and case reports.

INDEXING/ABSTRACTING

The *WJCC* is now indexed in Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports/Science Edition, Scopus, PubMed, and PubMed Central. The 2021 Edition of Journal Citation Reports® cites the 2020 impact factor (IF) for *WJCC* as 1.337; IF without journal self cites: 1.301; 5-year IF: 1.742; Journal Citation Indicator: 0.33; Ranking: 119 among 169 journals in medicine, general and internal; and Quartile category: Q3. The *WJCC*'s CiteScore for 2020 is 0.8 and Scopus CiteScore rank 2020: General Medicine is 493/793.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Ying-Yi Yuan*, Production Department Director: *Xu Guo*, Editorial Office Director: *Jin-Lei Wang*.

NAME OF JOURNAL

World Journal of Clinical Cases

ISSN

ISSN 2307-8960 (online)

LAUNCH DATE

April 16, 2013

FREQUENCY

Thrice Monthly

EDITORS-IN-CHIEF

Bao-Gan Peng, Jerzy Tadeusz Chudek, George Kontogeorgos, Maurizio Serati, Ja Hyeon Ku

EDITORIAL BOARD MEMBERS

<https://www.wjgnet.com/2307-8960/editorialboard.htm>

PUBLICATION DATE

June 26, 2022

COPYRIGHT

© 2022 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>

Clinical and Translational Research

Construction and validation of a novel prediction system for detection of overall survival in lung cancer patients

Cheng Zhong, Yun Liang, Qun Wang, Hao-Wei Tan, Yan Liang

Specialty type: Oncology**Provenance and peer review:**

Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind**Peer-review report's scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B

Grade C (Good): C

Grade D (Fair): 0

Grade E (Poor): 0

P-Reviewer: Miranda LA, Brazil; Mizuno N, Japan**Received:** February 21, 2022**Peer-review started:** February 21, 2022**First decision:** March 23, 2022**Revised:** March 30, 2022**Accepted:** April 29, 2022**Article in press:** April 29, 2022**Published online:** June 26, 2022**Cheng Zhong, Qun Wang, Hao-Wei Tan**, Department of Respiratory, Fenghua District People's Hospital, Ningbo 315000, Zhejiang Province, China**Yun Liang, Yan Liang**, Department of Hematology and Oncology, Fengdu People's Hospital, Chongqing 408200, China**Corresponding author:** Yun Liang, MD, Attending Doctor, Department of Hematology and Oncology, Fengdu People's Hospital, No. 33 Lutang Street, Sanhe Town, Chongqing 408200, China. dr_ly123@163.com

Abstract

BACKGROUND

Many factors have an aberrant effect on the overall survival of lung cancer (LC) patients. In recent years, remarkable progress has been made in immunotherapy, targeted treatment, and promising biomarkers. However, the available treatments and diagnostic methods are not specific for all patients.

AIM

To establish a system for predicting poor survival in patients with LC.

METHODS

The expression matrix and clinical information for this study were obtained from The Cancer Genome Atlas and Gene Expression Omnibus databases. After the differential analysis of all screened genes, weighted gene coexpression network analysis was performed to analyze hub genes related to patient survival. A logistic regression model was used to construct the scoring system. The expression of the hub genes was verified by performing quantitative reverse transcription-polymerase chain reaction.

RESULTS

A total of 5007 differentially expressed genes were selected for the Weighted Gene Co-expression Network Analysis algorithm. We found that the turquoise module showed the highest correlation with patient prognosis. The gene module with the greatest positive correlation with patient survival was located in the turquoise area. The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses performed for the genes contained in the turquoise module indicated the potential roles of the selected genes in the regulation of LC development. In addition, protein-protein interaction analysis was performed to screen hub genes,

which identified 100 hub genes located in the core area of the network. We then intersected the 100 hub genes with 75 key genes sorted by module members to identify real hub genes associated with prognosis. Forty-one genes were finally selected. We then used a logistic regression model to determine 11 independent risk genes, namely *CCNB2*, *CDC20*, *CENPO*, *FOXM1*, *HJURP*, *NEK2*, *OIP5*, *PLK1*, *PRC1*, *SKA1*, *UBE2C* and *SPARC*.

CONCLUSION

We constructed a predictive model based on 11 independent risk genes to establish a system predicting the survival status of patients with non-small-cell lung carcinoma.

Key Words: Lung cancer; Weighted Gene Co-expression Network Analysis; Hub genes; prognosis; Logistic regression

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: This was a bioinformatics-based study aimed at identifying a novel system for predicting overall survival in lung cancer patients. We constructed a predictive model using Weighted Gene Co-expression Network Analysis, protein-protein interaction network, and least absolute contraction and selection operator-logistic regression analysis. And the expression of hub genes was verified by polymerase chain reaction, immunohistochemistry in lung cancer cell lines, and patient samples.

Citation: Zhong C, Liang Y, Wang Q, Tan HW, Liang Y. Construction and validation of a novel prediction system for detection of overall survival in lung cancer patients. *World J Clin Cases* 2022; 10(18): 5984-6000

URL: <https://www.wjgnet.com/2307-8960/full/v10/i18/5984.htm>

DOI: <https://dx.doi.org/10.12998/wjcc.v10.i18.5984>

INTRODUCTION

Lung cancer (LC) is one of the most common malignant tumors and one of the leading causes of cancer-related deaths worldwide. In 2012, the deaths caused by LC were approximately 1.6 million, accounting for 19% of the total global cancer deaths[1,2]. Despite advancement in its treatment, surgery is the primary therapy for patients with non-small-cell lung carcinoma. However, the overall survival rate of LC patients remains low.

Many factors have an aberrant effect on the overall survival of LC patients. The main reason is that patients who are diagnosed with advanced and metastasis LC cannot undergo radical surgery. Therefore, the development of more advanced diagnosis and predictive biomarkers is a promising direction for cancer diagnosis and treatment[3,4].

In recent years, remarkable progress has been made in immunotherapy, targeted treatment, and promising biomarkers. However, the available treatments and diagnostic methods are not specific for all patients[5]. A high recurrence rate is observed after such treatment because of the complexity of cancer. Identification of new diagnostic and therapeutic biomarkers for cancer treatment is urgent[6].

The development of high-throughput technology has made important contributions to the identification of a large number of target genes in various diseases[7]. At the same time, as an emerging cross-discipline, bioinformatics analysis is widely used in the discovery of disease-related genes, new drug molecular targets, drug design, and functional analysis, which is helpful for the discovery of disease mechanisms[8]. Xie *et al*[9] performed bioinformatics analysis to analyze tumorigenesis-related genes and their target miRNAs in colon cancer, which facilitated the exploration of the potential targets for diagnosis, prognosis and treatment of colon carcinoma. Using RNA-Seq and bioinformatics methods, several key genes including *ID1*, *ID3* and *SMAD9* were identified in esophageal squamous cell carcinoma[10]. Many genes associated with LC progression and invasion have been identified by a combination of bioinformatics analysis and high-throughput sequencing[11-13].

The identification of differentially expressed genes (DEGs) has garnered considerable scientific attention. However, this method does not consider genes with similar expression patterns. Weighted Gene Co-expression Network Analysis (WGCNA) is a new algorithm that evaluates the correlation between gene modules and clinical features by constructing a scale-free gene coexpression network. In this study, we combined the WGCNA algorithm with DEGs to identify pivotal genes associated with clinicopathological characteristics and to provide insights into targeted therapy of LC.

MATERIALS AND METHODS

Data collection

The clinical and expression data of LC patients were derived from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases (<https://portal.gdc.cancer.gov/>; <http://www.ncbi.nlm.nih.gov/geo/>). GEO data contains two cohorts (GSE30129 and GSE50081). The *sva* package was used to normalize the Meta-GEO data. Next, we used the TCGA data ($\text{LogFC} > 0.5$, $P < 0.05$) to identify the DEGs and combined these DEGs with all GEO genes. Finally, 5007 genes were selected for the subsequent analyses.

Construction of WGCNA

WGCNA R package was used to analyze the coexpression networks. We determined the threshold of $\beta = 5$ to establish the optimal weighted network by Pearson's correlational analysis. The adjacent matrix was transformed into a topological overlap measure matrix *via* topological overlapping dissimilarity to estimate its connectivity property in the network. We set the minimum number of module genes to 100, and the threshold for merging similar modules was set to 0.25. $P < 0.05$ was considered to indicate statistical significance. After the modules of interest were selected, the key genes were selected according to the gene signature (GS) and module membership (MM) of each module.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses

The functional analysis of core genes was performed using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Through the R language clusterProfiler and ggplot2 packages, several important pathways have been discovered so far. The cut-off criteria were defined as count > 2 and $P < 0.05$.

Protein–protein interaction network

We employed the STRING database to analyze the interaction between the module genes and set the confidence score to ≥ 0.9 . The Cytoscape plug-in of Cytoscape software (version 3.7.0) was used to identify the core genes in the network.

Construction of predictive model

All patients were assigned to training and validation sets in the ratio of 6:4. The least absolute contraction and selection operator (LASSO) reduced the data dimensionality, and Cox regression analysis was applied to construct a patient prognostic evaluation model. The predictive efficacy of the model was evaluated by the receiver operating characteristic (ROC) curve. A nomogram was used to visualize the scoring system through the *rms* package in the R software.

Cell culture

A549 and H1299 LC cells, and human lung fibroblasts were purchased from the Cancer Cell Repository (Shanghai Cell Bank, Shanghai, China). The medium used for cell culture was 10% Dulbecco's modified eagle's medium (supplemented with fetal bovine serum and penicillin/streptomycin). The cells were cultured in an incubator under 5% CO₂ at 37 °C.

Quantitative real-time polymerase chain reaction

We use 1 mL TRIzol (Invitrogen, Grand Island, NY, United States) and 200 μL chloroform to extract the total RNA from 2×10^6 cells in LC cells. Total RNA was reverse transcribed into cDNA (TaKaRa Bio, Shiga, Japan). The cDNA, primers, and the SYBR Green PCR Master Mix (TOYOBO, Osaka, Japan) were quantitatively detected by PCR. The primer sequences of all genes are depicted in [Table 1](#). The gene expression level was evaluated by the $2^{-\Delta\Delta\text{Ct}}$ method. All experiments were repeated three times.

Statistical analyses

All data in this study were analyzed using GraphPad Prism 5 and R software. The data were expressed as mean \pm SD. A two-tailed *t* test was applied for quantitative real-time polymerase chain reaction (qRT-PCR) analysis among different groups. The results were considered to be significant at $P < 0.05$.

RESULTS

Identification of intersecting genes between GEO cohort and TCGA dataset

We screened DEGs based on the TCGA dataset by including $1037^{\text{tumor}}/108^{\text{normal}}$ samples. A total of 10 970 DEGs were selected based on the criteria of $P < 0.05$ and $|\log_2 \text{FC}| > 1$. The top 30 up- and downregulated genes are shown in [Figure 1A](#). We then intersected DEGs of TCGA with all genes in the GEO dataset and found 5007 common genes for further WGCNA.

Table 1 Sequence of polymerase chain reaction primers used in this study

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
CCNB2	CCGACGGTGTCCAGTGATTT	TGTGTTTTGGTGGGTGAACCT
CDC20	GCACAGTTCGCGTTCCGAGA	CTGGATTGCCAGGAGTTCGG
CENPO	AGTGAGCAGATCCCGTAAACA	GGTTGGGTCTACATTGGCAATA
FOXM1	CGTCGGCCACTGATTCTCAA	GGCAGGGGATCTCTTAGGTTCC
HJURP	CCACGCTGACCTACGAGAC	CTCACCGCTTTTGAATCGGC
GADPH	ACAACITTTGGTATCGTGGAAAGG	GCCATCACGCCACAGTTTC
NEK2	TGCTTCGTGAACTGAAACATCC	CCAGAGTCAACTGAGTCATCACT
OIP5	TGAGAGGGCGATTGACCAAG	AGCACTGCGTGACACTGTG
PLK1	AAAGAGATCCCGGAGGTCTCA	GGCTGCGGTGAATGGATATTTCC
PRC1	ATCACCTTCGGGAAATATGGGA	TCTTTCTGACAGACGGATATGCT
SKA1	CCTGAACCCGTAAGAAGCCT	TCATGTACGAAGGAACACCATTG
UBE2C	GACCTGAGGTATAAGCTCTCGC	TTACCTGGGTGTCCACGTT

WGCNA

To determine the roles of common DEGs associated with prognosis and other clinicopathological characteristics of LC patients, WGCNA was performed to construct a coexpression network. As shown in [Figure 1B](#) and [C](#), the correlation coefficient was converted to the adjacent coefficient according to the optimal parameter ($\beta = 5$). Thereafter, highly correlated samples and delete discrete samples were clustered ([Figure 1E](#)). A threshold of 0.25 and a minimum gene number of 150 were considered to merge similar modules. [Figure 1D](#) shows eight modules that were finally selected on the basis of the filter criteria. The hierarchical clustering of module hub genes is shown in [Figure 1F](#).

Identification of highly correlated modules

The topological overlap matrix plot ([Figure 1G](#)) indicated the correlation between the genes of the eight modules sorted using the clustering tree. The turquoise module showed a positive correlation of about 0.31 with LC patient survival, followed by the green module (0.28) and yellow module (0.23) ([Figure 2A](#) and [B](#)). The turquoise module contained 1673 genes. We then selected 75 key genes from the turquoise module with $MM > 0.8$ ([Figure 2C](#)). Taken together, the turquoise module was finally selected for further analysis.

GO and KEGG analyses in modules

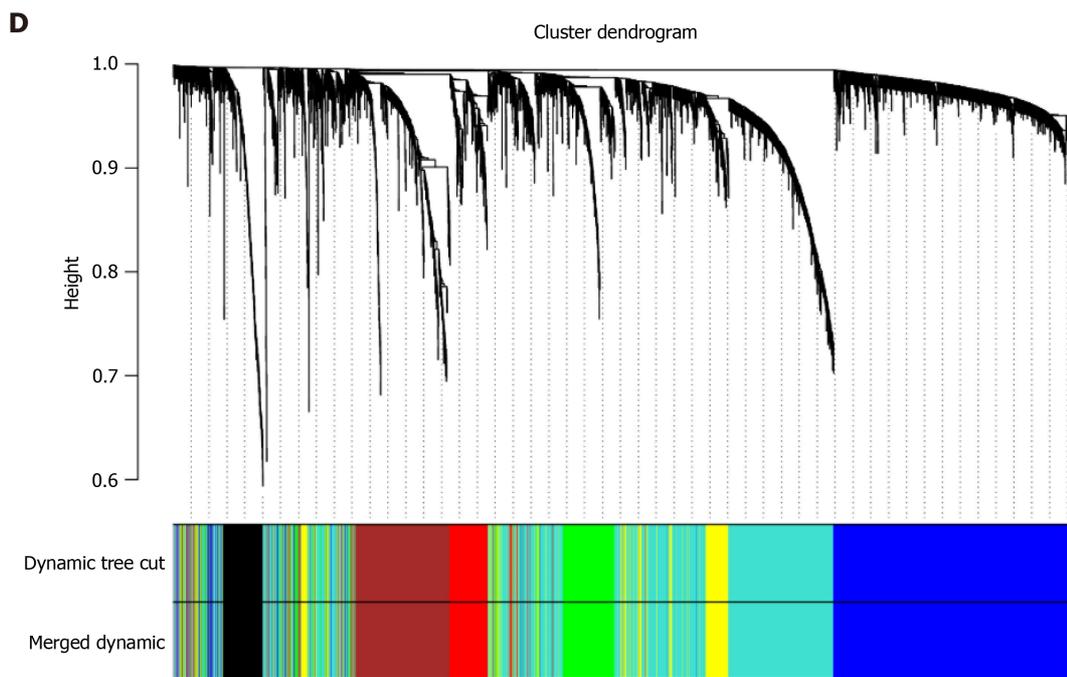
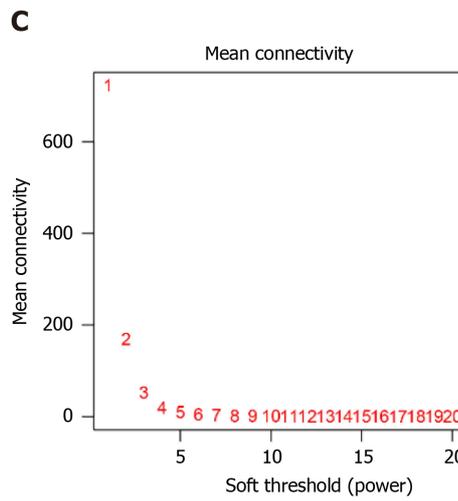
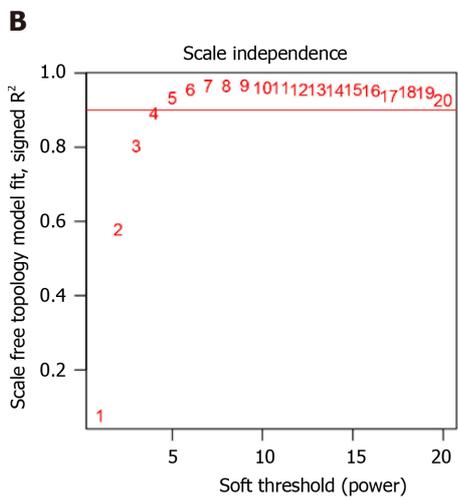
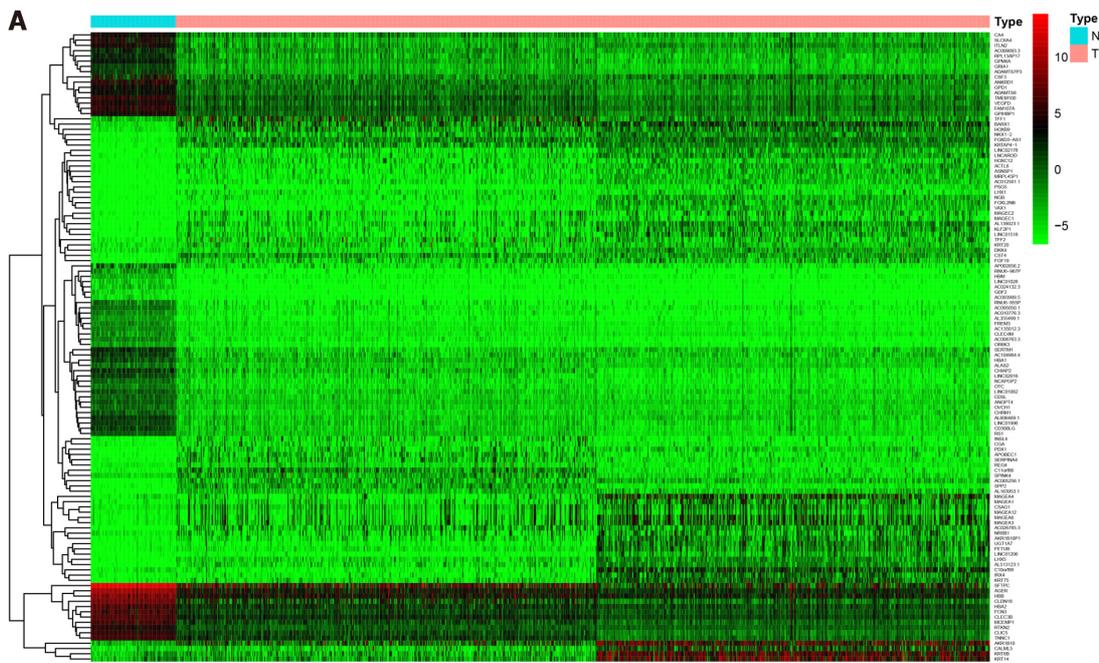
GO enrichment analysis experiments showed that the turquoise module genes mainly encoded for ATPase, helicase, 3'-5' DNA helicase, DNA-dependent ATPase, DNA helicase, ATP-dependent DNA helicase, and ATP-dependent helicase and associated with the binding of many molecules including DNA replication origin, single-stranded DNA, and tubulin. KEGG analysis indicated that many signaling pathways involved in Fanconi anemia and the p53 signaling pathway were correlated with turquoise module genes. In addition, other important pathways such as metabolism of carbon, pyrimidine, cysteine, and methionine, cell cycle, and DNA repair were also found in the turquoise module. These results indicated that the mechanism that affects the survival of LC patients may be closely related to the molecular binding mechanism and several important signaling pathways ([Figure 2D](#)).

Establishment of protein-protein interaction networks and selection of module genes

To determine which cluster of genes in the turquoise module have a pivotal effect on the prognosis of LC, we constructed protein-protein interaction networks using the STRING database ([Figure 3A](#)) and Cytoscape software and found 100 hub genes located in the core area of the network ([Figure 3B](#)). We then intersected the 100 hub genes with 75 key genes sorted by MM to identify real hub genes associated with prognosis ([Figure 3C](#)).

Construction of the hub-genes-based scoring system

Subsequently, we performed a LASSO-logistic analysis of real hub genes to establish a prognostic evaluation model. Finally, 11 prognostic genes were selected in the predictive model in the training dataset, namely *CCNB2*, *CDC20*, *CENPO*, *FOXM1*, *HJURP*, *NEK2*, *OIP5*, *PLK1*, *PRC1*, *SKA1*, *UBE2C* and *SPARC* ([Figure 4A](#)). The risk score = 4.43 (Intercept) + *CCNB2*-expression \times 0.552 + *CDC20*-expression \times 0.037 + *CENPO*-expression \times 0.287 + *FOXM1*-expression \times 0.106 + *HJURP*-expression \times 0.229 + *NEK2*-



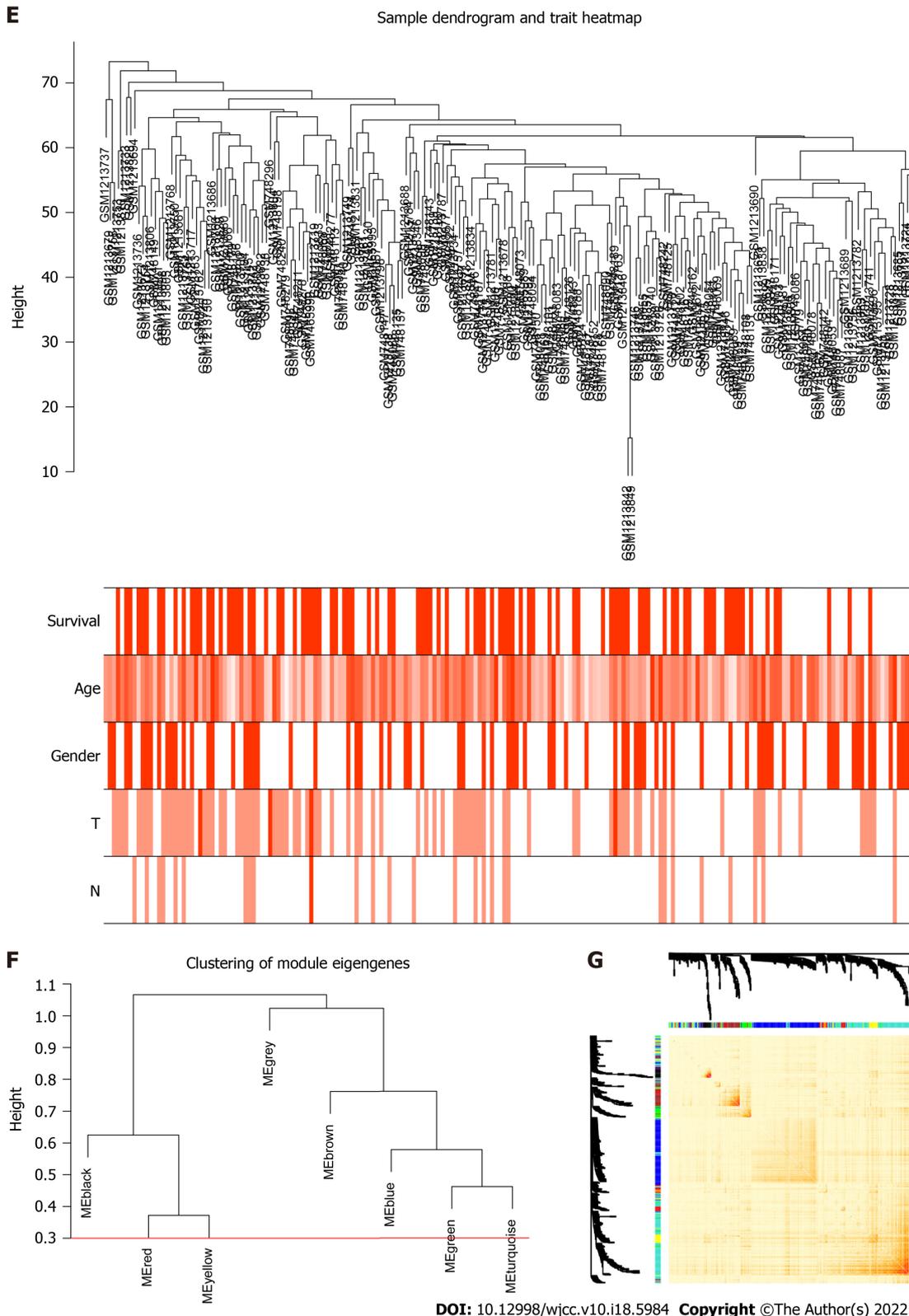
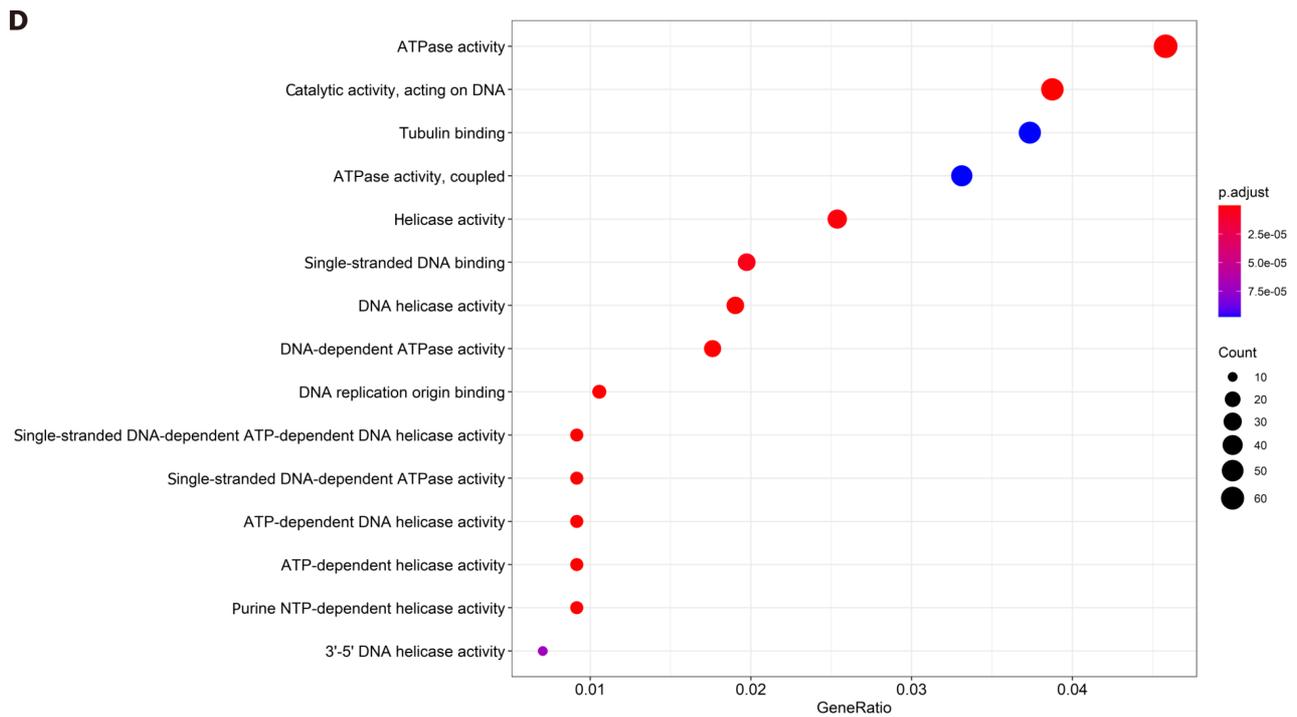
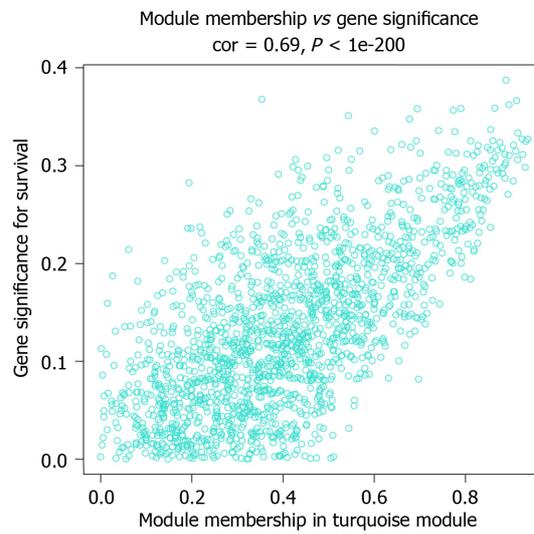
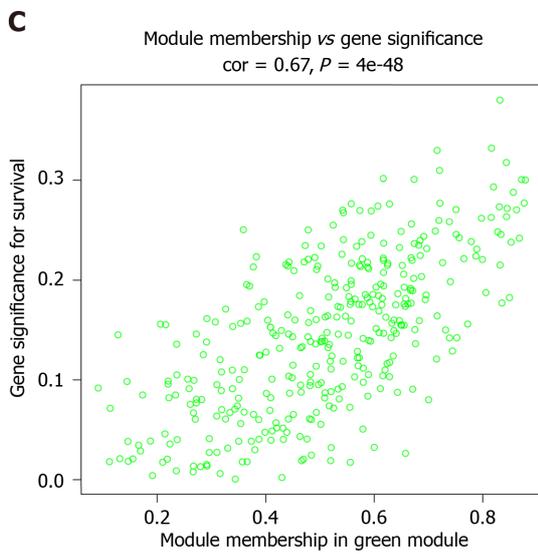
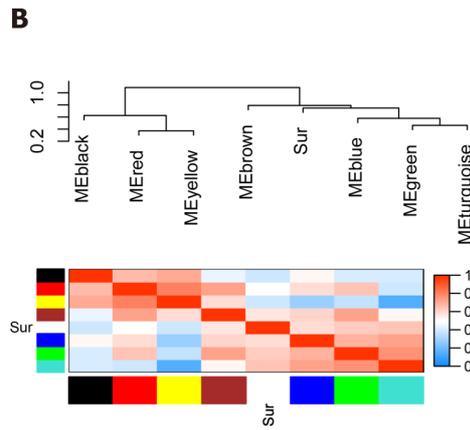
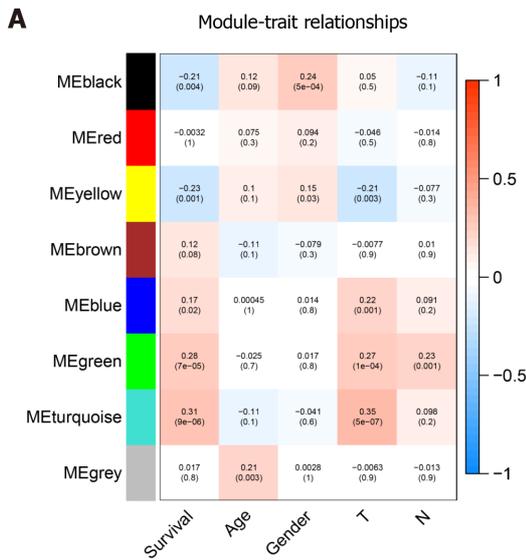


Figure 1 Identification of differentially expressed genes and Weighted Gene Co-expression Network Analysis. A: The differentially expressed genes analyzed in The Cancer Genome Atlas dataset. Top 30 upregulated and downregulated genes are shown; B and C: Soft-threshold power analysis revealed the scale-free fit index and the mean connectivity of network topology; D: Hierarchical cluster analysis of the coexpression module based on the dissimilarity measurement; E: Sample clustering based on expression data used to detect the outliers; F: Dendrogram of consensus module eigen genes. Groups of eigen genes below the red line merged owing to their similarity; G: The topological overlap matrix heatmap showing the overlap between the co-expression genes.

expression $\times 0.083$ OIP5-expression $\times 0.020$ PLK1-expression $\times 0.520$ + PRC1-expression $\times 0.192$ SKA1-expression $\times 0.110$ + UBE2C-expression $\times 0.263$. Subsequently, we evaluated the reliability of the model by the ROC curve. The results showed that the area under the curve of the training set and test set were 0.754 and 0.626, respectively (Figure 4B). Cox regression analysis showed that risk score was an



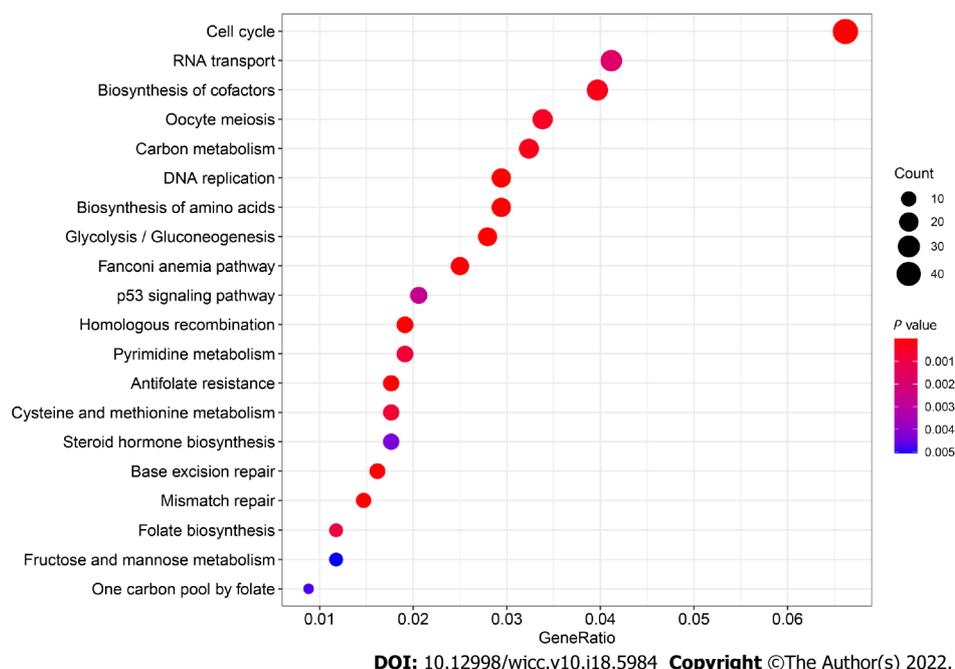


Figure 2 Identification of highly correlated gene modules by Weighted Gene Co-expression Network Analysis. A: Correlation between eight module genes and the clinical features. Turquoise module indicating a high correlation with the patient's overall survival ($P < 0.001$, $r = 0.31$); B: Heatmap plot showing the adjacent modules and survival traits; C: The gene signature and module membership of turquoise and green modules; D: Significantly enriched Gene Ontology items and the Kyoto Encyclopedia of Genes and Genomes pathways in turquoise module with top 20 count number of genes shown.

independent risk factor for predicting the poor prognosis of LC patients (Figure 4C). To further evaluate the prognosis of LC, we constructed a nomogram based on risk factors (Figure 4D and E).

After the construction of the scoring system in the GEO dataset, we determined the effect of 11 genes in the TCGA dataset. As shown in Figure 5A, all genes were differentially expressed in LC patients compared with normal samples. Moreover, all genes were significantly correlated with patient prognosis except for *CCNB2* and *SKA1* (Figure 5B).

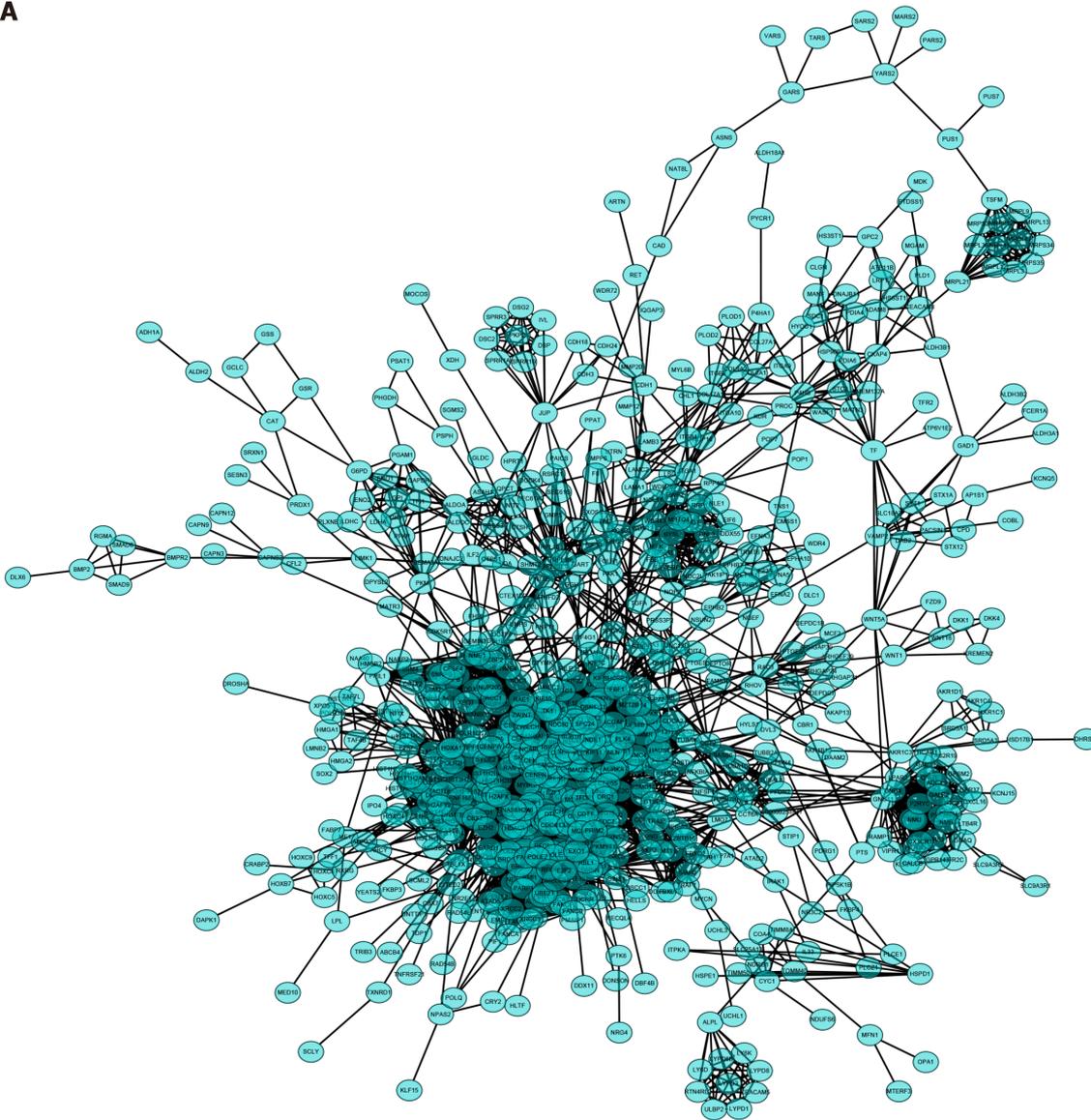
To validate the expression of the 11 hub genes, we performed an immunohistochemistry experiment obtained from the Protein Atlas database. Immunohistochemistry indicated that the protein levels of *CDC20*, *FOXM1*, *HJURP*, *PRC1*, *UBE2C* and *CCNB2* were increased in the LC samples compared with the normal samples, whereas those of *OIP5*, *PLK1* and *SKA1* were decreased (Figure 5C). To explore the mRNA expression levels of these genes, we performed qRT-PCR analysis and found that the mRNA levels of *CCNB2*, *CDC20*, *FOXM1*, *HJURP*, *NEK2*, *PRC1*, *SKA1* and *UBE2C* were increased in the LC cell lines, whereas those of *CENPO*, *OIP5* and *PLK1* was decreased (Figure 5D).

DISCUSSION

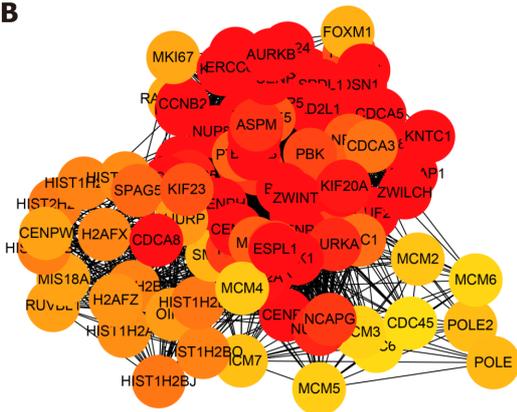
LC treatments include a combination of radical surgery, radiation therapy, chemotherapy, and precise targeted therapy[14]. Despite advancement in LC treatment and diagnosis, the 5-year overall survival rate remains low[15]. The main reason is that patients who are diagnosed with advanced and metastasis LC cannot undergo radical surgery. Therefore, more specific and sensitive biomarkers are needed to facilitate early diagnosis and prediction of overall survival.

Recently, several therapeutic targets and prognostic biomarkers have been identified using advanced high-throughput sequencing technology and integrative bioinformatics analysis. Previous studies have reported many prognostic biomarkers in LC by performing a combined analysis using TCGA and GEO datasets and validated by *in vitro* experiments. Sun *et al*[16] reported the role of C-type lectin domain family 3 member B (CLEC3B) in tumor progression, prognosis, and immune responses in LC by performing RNA-Seq and bioinformatics analysis; the expression and methylation of CLEC3B were also validated by qRT-PCR analysis. miRNA-144-3p, an important noncoding RNA, was identified and validated as an independent risk factor for LC prognosis by performing bioinformatics analysis and qRT-PCR[17]. However, because of the insufficient sample size, biological heterogeneity, and different statistical methods, highly effective genes are not found in clinical practice. Moreover, the prediction efficiency in tumor patients could be limited by simply using a single GS instead of a multi-GS. Therefore, more biological markers and more effective prediction models are required for the prevention and treatment of LC.

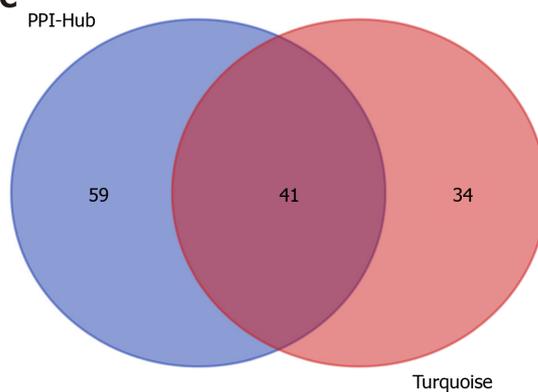
A



B



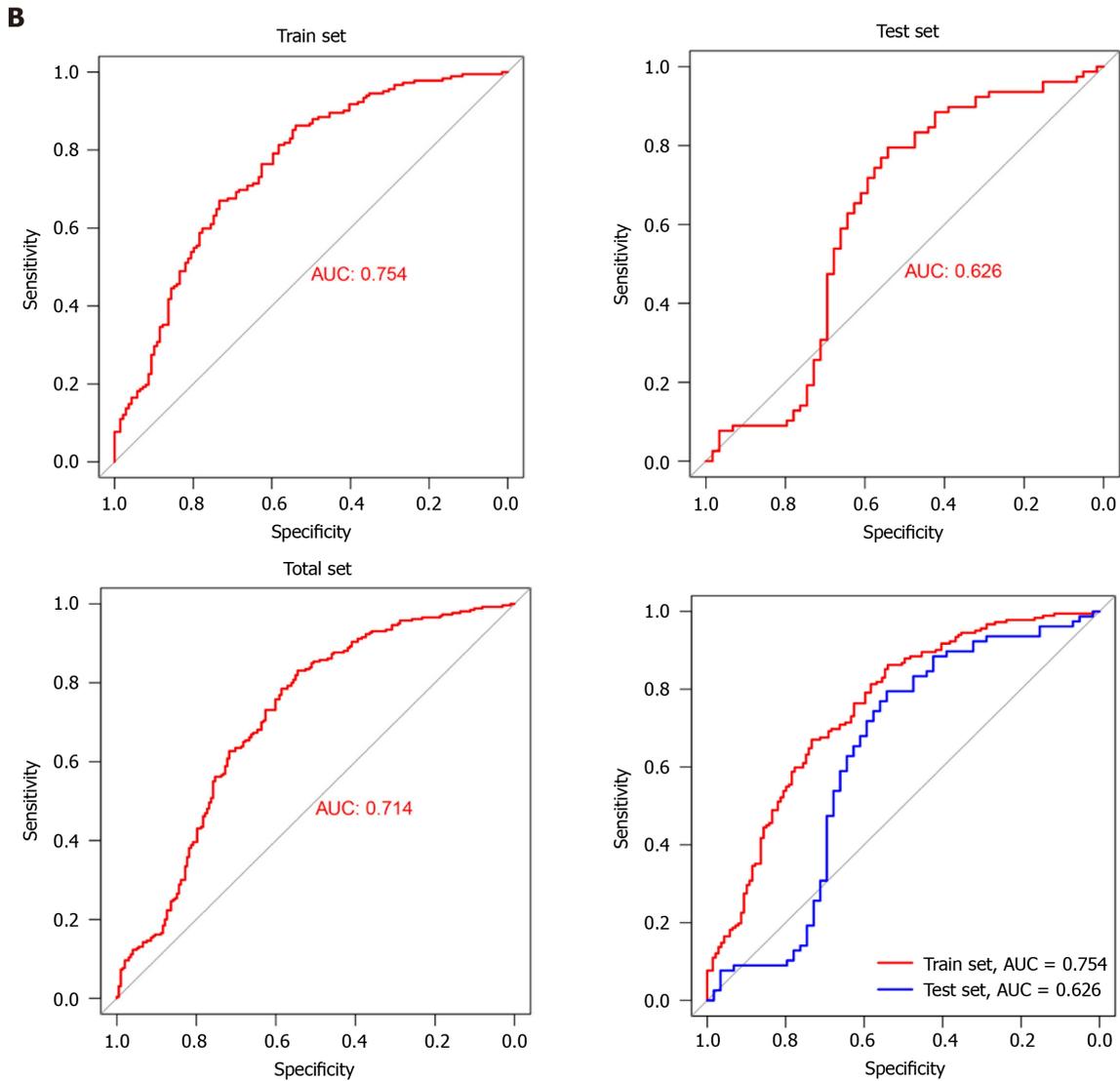
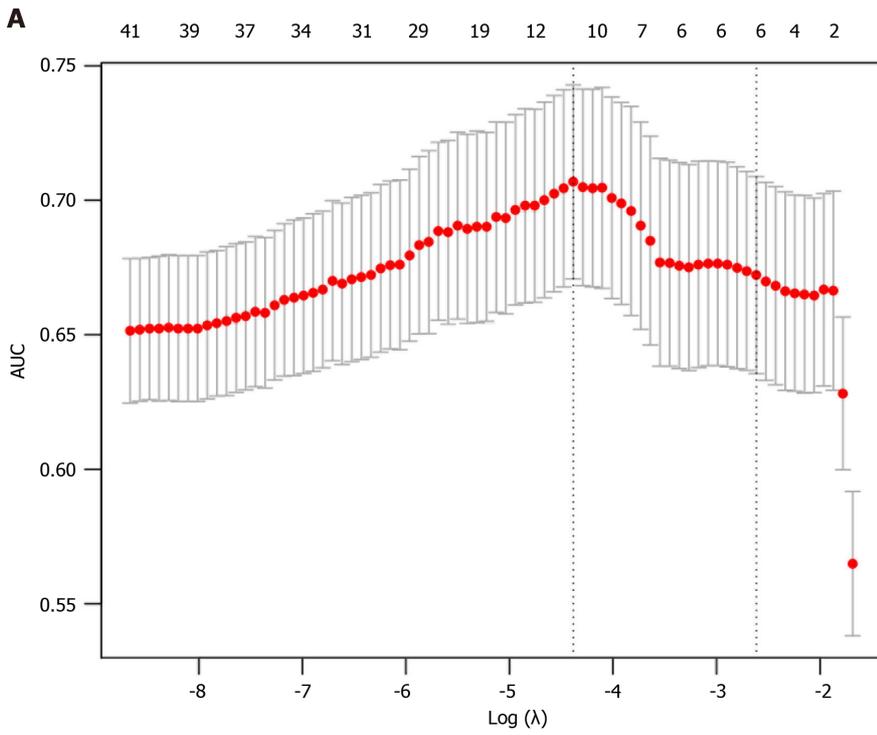
C



DOI: 10.12998/wjcc.v10.i18.5984 Copyright ©The Author(s) 2022.

Figure 3 Construction of the protein–protein interaction networks of turquoise module genes and the selected hub genes. A: Network of all turquoise module genes excluding the low connectivity genes; B: The cytohubba algorithm used to identify the top 100 hub genes located in the core area of the turquoise module; C: Venn diagram showing an overlap between the protein–protein interaction hub and gene signature/module membership-key genes in the turquoise module. A total of 41 real hub genes finally selected for further analyses.

In the present study, we initially detected DEGs based on the TCGA dataset and intersected these DEGs with all GEO cohort genes to obtain an expression profile. We used the WGCNA algorithm to identify core genes in GEO expression data and that were highly related to clinical features. WGCNA



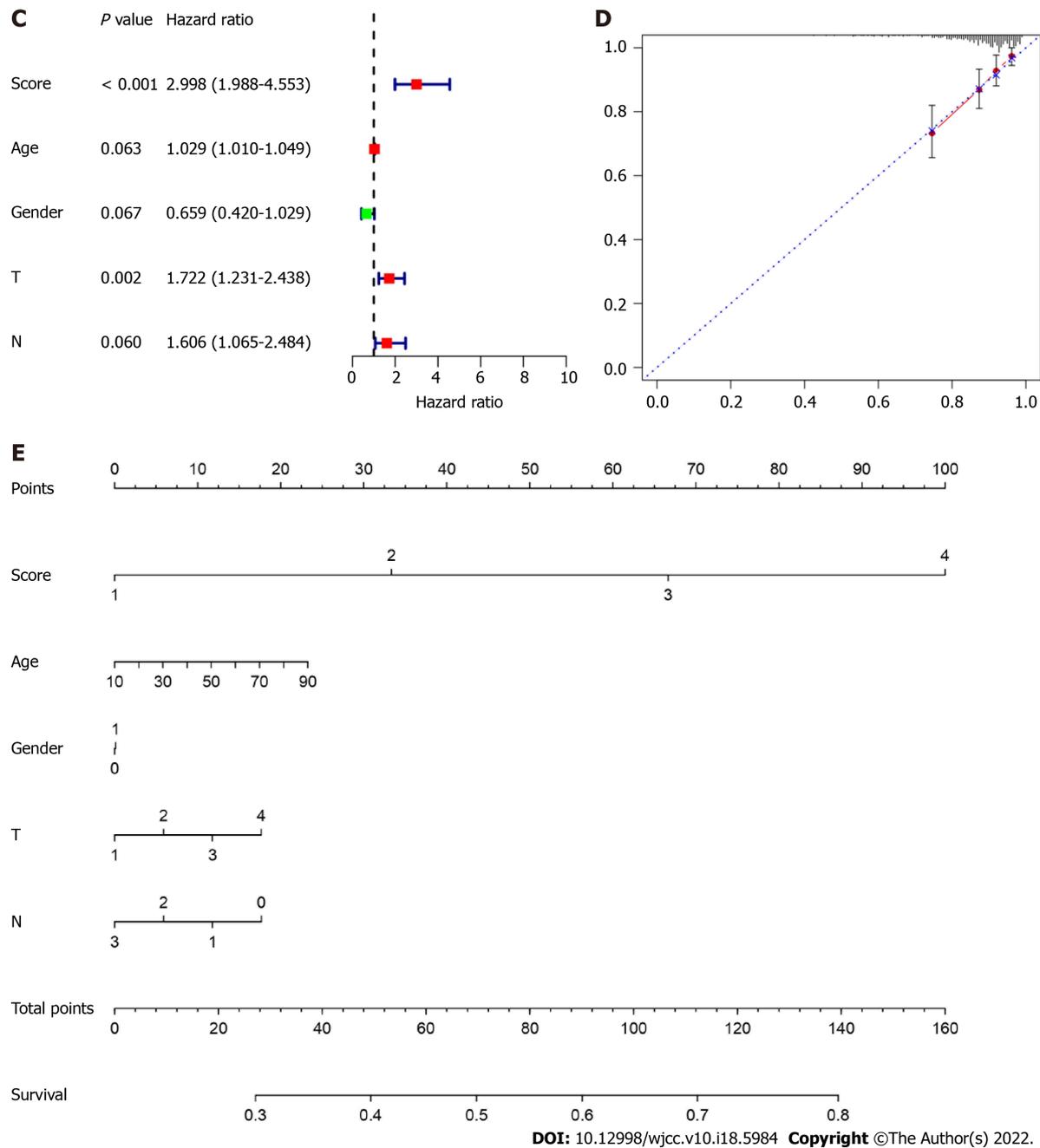
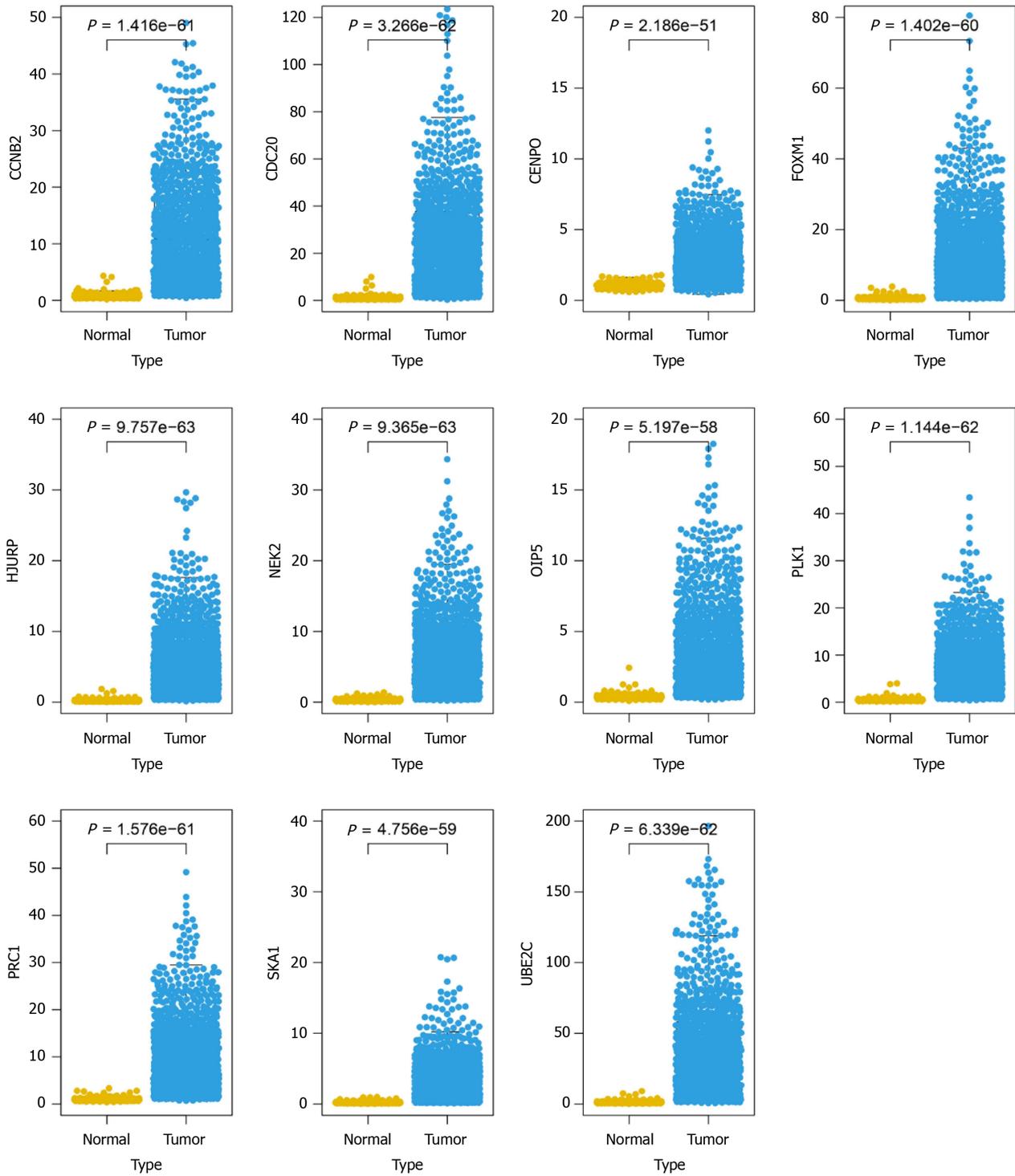


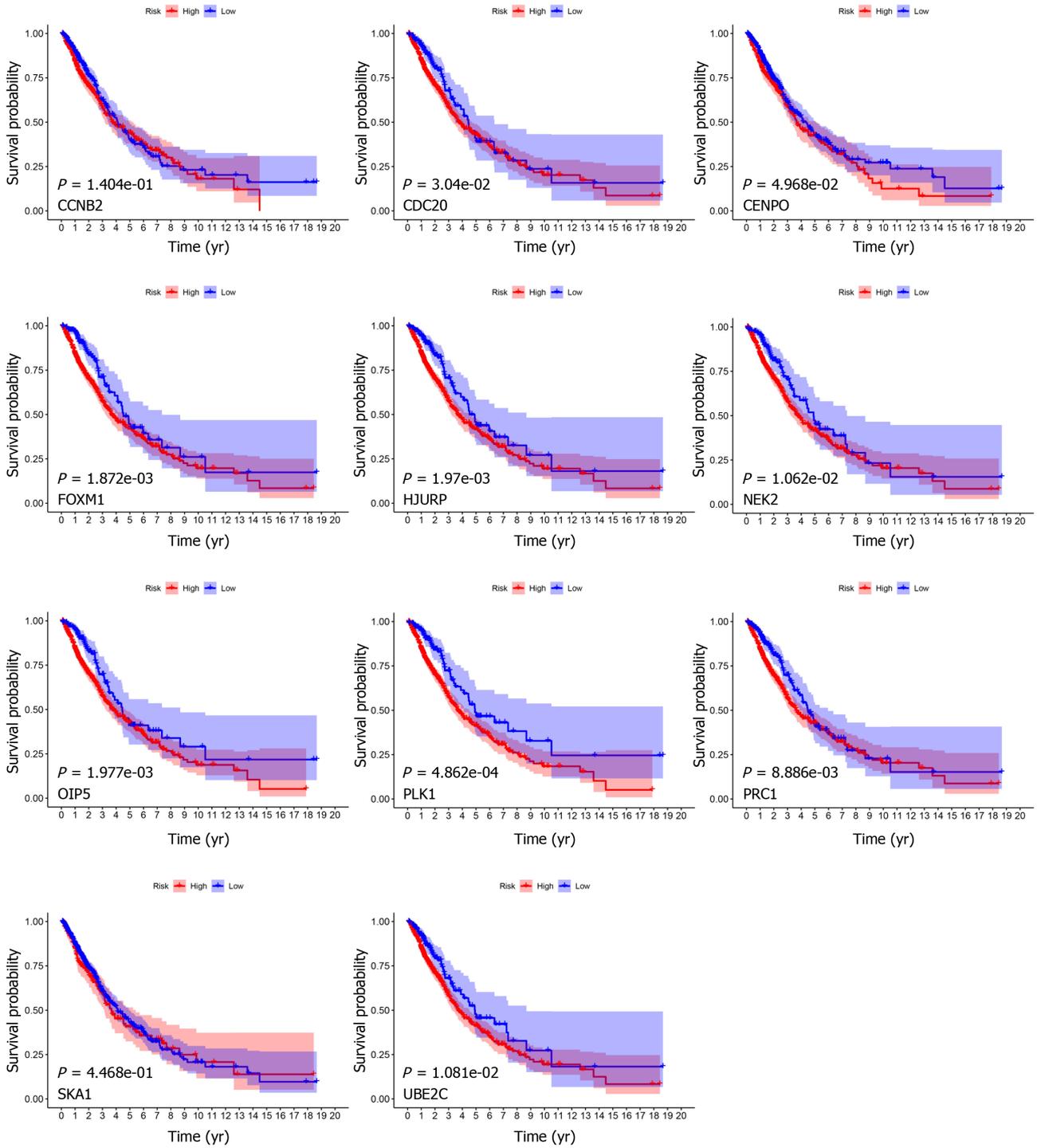
Figure 4 Construction of a prognostic predictive model using real hub genes. A: Area under the curve measurement performed to select the optimal candidate genes for the scoring system construction; B: Receiver operating characteristic curve showing the prediction efficiency of the scoring system in the training, test, and all dataset; C: Risk-score served as an independent risk factor in predicting patients' survival; D and E: Development of nomograms and calibration curves for prognosis in the total dataset.

classified eight modules and subsequently correlated the modules with clinical characteristics. In these modules, the turquoise module contained 1673 genes that showed the highest correlation with LC patient prognosis. We then performed GO enrichment and KEGG pathway analyses of the turquoise module genes and found that the function of these genes was mainly related to activation of enzymes including ATPase, helicase, 3'-5' DNA helicase, DNA-dependent ATPase, DNA helicase, ATP-dependent DNA helicase, and ATP-dependent helicase and binding of many molecules including DNA replication origin, single-stranded DNA, and tubulin and activation of signaling pathways involved in Fanconi anemia and p53 signaling pathway. Subsequently, we performed a protein-protein interaction network analysis of the genes contained in the yellow module and intersected the network hub genes with $MM > 0.8$. Forty-one genes were selected and subjected to LASSO-logistic regression. We finally identified 11 prognostic genes, namely *CCNB2*, *CDC20*, *CENPO*, *FOXM1*, *HJURP*, *NEK2*, *OIP5*, *PLK1*, *PRC1*, *SKA1*, *UBE2C* and *SPARC*. Among these genes, *FOXM1* and *PLK1* are the most studied genes in LC. *FOXM1*, an important family member of the FOX family, plays a pivotal role in a series of biological processes, including facilitating cell proliferation, differentiation, and organ development[18]. *FOXM1*

A



B



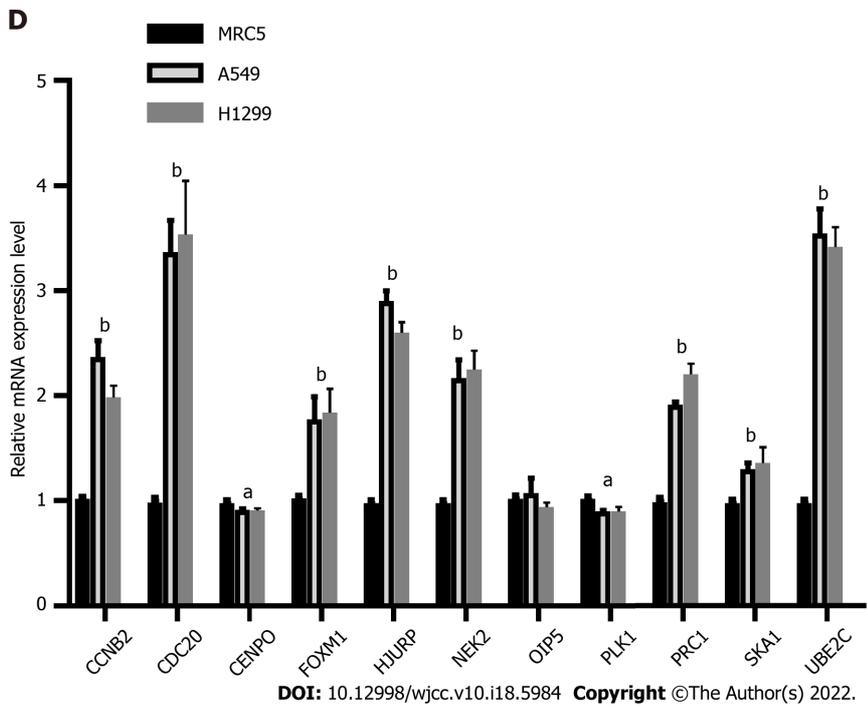
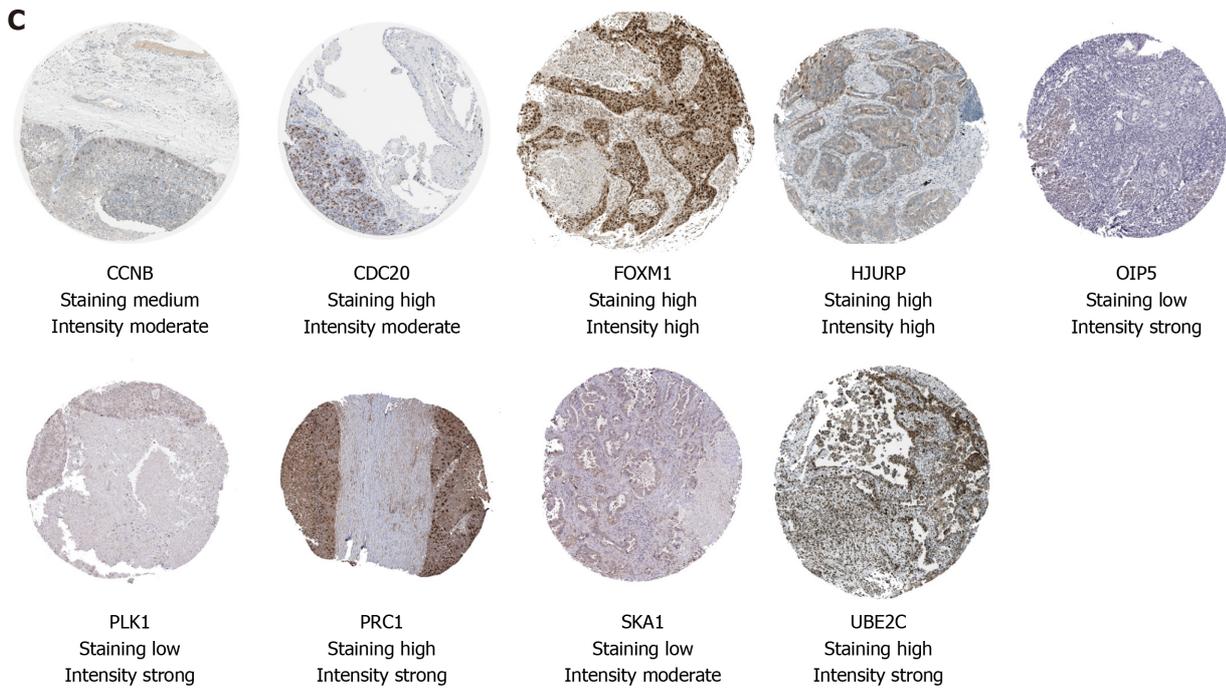


Figure 5 Expression and survival curve of 11 genes in The Cancer Genome Atlas dataset and the validation of the 11 genes' expression by immunohistochemistry and quantitative real-time polymerase chain reaction analyses. A: Expression of 11 genes in The Cancer Genome Atlas (TCGA) dataset; B: Correlation of 11 genes and the survival rate in the TCGA dataset; C: Expression profile of 11 genes in tumor tissues obtained from the Protein Atlas database; D: Quantitative real-time polymerase chain reaction analysis indicating the expression of 11 genes in lung cancer cell lines when compared with that in normal lung bronchial cells. ^a*P* < 0.05, ^b*P* < 0.01.

level is significantly increased in LC cells and could be regulated by miR-216b, which promotes cancer progression and epithelial-mesenchymal transition in LC cells[19]. Moreover, FOXM1 could directly regulate the radiosensitivity of LC cells *via* interacting with KIF20A, suggesting that FOXM1 might be a novel therapeutic target for LC treatment. FOXM1 could also be regulated by other important molecules. The family with sequence similarity 188-member B is a member of the novel putative deubiquitinase family and directly binds to FOXM1, which promotes LC progression[20]. PLK1 is highly correlated with LC progression. PLK1 can target and regulate the transforming growth factor β signaling pathway, and then amplify its metastatic activity by positive feedback[20]. PLK1 is also regulated by long noncoding RNAs. For instance, miR-296-5p decreases the ability of cell invasion and migration by directly targeting PLK1 in LC cells[21]. However, other genes have not been actively

researched, especially with regard to the mechanism of progression in LC. Therefore, in-depth knowledge of these genes will help develop new biomarkers for early LC diagnosis and prediction of prognosis.

After the scoring system was constructed, we further evaluated the performance of the model in LC patients. The ROC curve showed that the model had an excellent predictive performance. In addition, the risk predictor of the model can be considered an independent risk factor for predicting LC prognosis. To conclude, this study showed the potential of prognostic genes in LC patients using WGCNA combined with the established predictive model. However, this study also had some limitations. Firstly, LC patients were from public databases, thus the number of samples was limited. In future studies, we will collect samples from our hospital to expand the sample size to validate the predictive model. Second, the molecular biological mechanism by which the hub gene affects the prognosis of patient needs to be further explored.

CONCLUSION

This study used the WGCNA algorithm to identify functional modules highly correlated with LC prognosis. After construction of the predictive model, we screened and validated 11 prognostic genes, which might be considered new therapeutic targets for the diagnosis and treatment of LC. This study also had some limitations. The mechanisms of the effect of the 11 prognostic genes on cancer progression need to be studied in the future.

ARTICLE HIGHLIGHTS

Research background

Many factors have an aberrant effect on the overall survival of lung cancer (LC) patients. In recent years, remarkable progress has been made in immunotherapy, targeted treatment, and promising biomarkers. However, the available treatments and diagnostic methods are not specific for all patients.

Research motivation

Identifying new diagnostic and therapeutic biomarkers for cancer treatment is urgent.

Research objectives

We aimed to establish a system for predicting poor survival in patients with LC.

Research methods

Weighted Gene Co-expression Network Analysis (WGCNA), functional enrichment analysis, quantitative real-time polymerase chain reaction, and other bioinformatics analysis were used in this study.

Research results

A total of 5007 differentially expressed genes were selected for the WGCNA algorithm. The turquoise module showed the highest correlation with patient prognosis. The gene module with the greatest positive correlation with patient survival was located in the turquoise area. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses performed for the genes contained in the turquoise module indicated the potential roles of the selected genes in the regulation of LC development. In addition, protein-protein interaction analysis was performed to screen hub genes, which identified 100 hub genes located in the core area of the network. We intersected the 100 hub genes with 75 key genes sorted by module members to identify real hub genes associated with prognosis. Forty-one genes were finally selected. We used a logistic regression model to determine 11 independent risk genes, namely *CCNB2*, *CDC20*, *CENPO*, *FOXM1*, *HJURP*, *NEK2*, *OIP5*, *PLK1*, *PRC1*, *SKA1*, *UBE2C* and *SPARC*.

Research conclusions

We constructed a model based on 11 independent risk genes to establish a system to predict the survival status of patients with non-small-cell lung carcinoma.

Research perspectives

The new predictive model could play a role in overall survival.

FOOTNOTES

Author contributions: Zhong C and Liang Y conceptualized and designed the article; Zhong C and Wang Q analyzed and interpreted the data; Zhong C drafted of the article; Liang Y and Tang HW were responsible for critical revision of the article for important intellectual content.

Institutional review board statement: This study was approved by the Ethics Committee of the Fenghua District People's Hospital.

Clinical trial registration statement: This study does not involve the clinical trials, so the clinical trial registration is not required.

Informed consent statement: The data that support the findings of current study are publicly available, so the signed informed consent document is not required.

Conflict-of-interest statement: The authors have no conflicts of interest to declare.

Data sharing statement: No additional data are available.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country/Territory of origin: China

ORCID number: Cheng Zhong 0000-0002-4077-2676; Yun Liang 0000-0003-3759-2997; Qun Wang 0000-0003-2258-4791; Hao-Wei Tan 0000-0001-6188-9294; Yan Liang 0000-0002-8608-0229.

S-Editor: Gong ZM

L-Editor: Kerr C

P-Editor: Gong ZM

REFERENCES

- 1 **Torre LA**, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65**: 87-108 [PMID: 25651787 DOI: 10.3322/caac.21262]
- 2 **Chen W**, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016; **66**: 115-132 [PMID: 26808342 DOI: 10.3322/caac.21338]
- 3 **Villalobos P**, Wistuba II. Lung Cancer Biomarkers. *Hematol Oncol Clin North Am* 2017; **31**: 13-29 [PMID: 27912828 DOI: 10.1016/j.hoc.2016.08.006]
- 4 **Di X**, Jin X, Li R, Zhao M, Wang K. CircRNAs and lung cancer: Biomarkers and master regulators. *Life Sci* 2019; **220**: 177-185 [PMID: 30711537 DOI: 10.1016/j.lfs.2019.01.055]
- 5 **Ciliberto D**, Staropoli N, Caglioti F, Gualtieri S, Fiorillo L, Chiellino S, De Angelis AM, Mendicino F, Botta C, Caraglia M, Tassone P, Tagliaferri P. A systematic review and meta-analysis of randomized trials on the role of targeted therapy in the management of advanced gastric cancer: Evidence does not translate? *Cancer Biol Ther* 2015; **16**: 1148-1159 [PMID: 26061272 DOI: 10.1080/15384047.2015.1056415]
- 6 **Tsoukalas N**, Kiakou M, Tampakidis K, Tolia M, Aravantinou-Fatorou E, Baxevanos P, Kyrgias G, Theocharis S. PD-1 and PD-L1 as immunotherapy targets and biomarkers in non-small cell lung cancer. *J BUON* 2019; **24**: 883-888 [PMID: 31424637]
- 7 **Moncho-Amor V**, Pintado-Berninches L, Ibañez de Cáceres I, Martín-Villar E, Quintanilla M, Chakravarty P, Cortes-Sempere M, Fernández-Varas B, Rodríguez-Antolín C, de Castro J, Sastre L, Perona R. Role of Dusp6 Phosphatase as a Tumor Suppressor in Non-Small Cell Lung Cancer. *Int J Mol Sci* 2019; **20** [PMID: 31027181 DOI: 10.3390/ijms20082036]
- 8 **Kulasingam V**, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* 2008; **5**: 588-599 [PMID: 18695711 DOI: 10.1038/ncponc1187]
- 9 **Xie B**, Zhao R, Bai B, Wu Y, Xu Y, Lu S, Fang Y, Wang Z, Maswikiti EP, Zhou X, Pan H, Han W. Identification of key tumorigenesis-related genes and their microRNAs in colon cancer. *Oncol Rep* 2018; **40**: 3551-3560 [PMID: 30272358 DOI: 10.3892/or.2018.6726]
- 10 **Zhao Y**, Zhu J, Shi B, Wang X, Lu Q, Li C, Chen H. The transcription factor LEF1 promotes tumorigenicity and activates the TGF- β signaling pathway in esophageal squamous cell carcinoma. *J Exp Clin Cancer Res* 2019; **38**: 304 [PMID: 31296250 DOI: 10.1186/s13046-019-1296-7]
- 11 **Zhao T**, Khadka VS, Deng Y. Identification of lncRNA biomarkers for lung cancer through integrative cross-platform data analyses. *Aging (Albany NY)* 2020; **12**: 14506-14527 [PMID: 32675385 DOI: 10.18632/aging.103496]
- 12 **Xue L**, Xie L, Song X. Identification of potential tumor-educated platelets RNA biomarkers in non-small-cell lung cancer by integrated bioinformatical analysis. *J Clin Lab Anal* 2018; **32**: e22450 [PMID: 29665143 DOI: 10.1002/jcla.22450]

- 13 **Liu L**, Ahmed T, Petty WJ, Grant S, Ruiz J, Lycan TW, Topaloglu U, Chou PC, Miller LD, Hawkins GA, Alexander-Miller MA, O'Neill SS, Powell BL, D'Agostino RB Jr, Munden RF, Pasche B, Zhang W. SMARCA4 mutations in KRAS-mutant lung adenocarcinoma: a multi-cohort analysis. *Mol Oncol* 2021; **15**: 462-472 [PMID: [33107184](#) DOI: [10.1002/1878-0261.12831](#)]
- 14 **Hensing T**, Chawla A, Batra R, Salgia R. A personalized treatment for lung cancer: molecular pathways, targeted therapies, and genomic characterization. *Adv Exp Med Biol* 2014; **799**: 85-117 [PMID: [24292963](#) DOI: [10.1007/978-1-4614-8778-4_5](#)]
- 15 **Li F**, He H, Qiu B, Ji Y, Sun K, Xue Q, Guo W, Wang D, Zhao J, Mao Y, Mu J, Gao S. Clinicopathological characteristics and prognosis of lung cancer in young patients aged 30 years and younger. *J Thorac Dis* 2019; **11**: 4282-4291 [PMID: [31737313](#) DOI: [10.21037/jtd.2019.09.60](#)]
- 16 **Sun J**, Xie T, Jamal M, Tu Z, Li X, Wu Y, Li J, Zhang Q, Huang X. CLEC3B as a potential diagnostic and prognostic biomarker in lung cancer and association with the immune microenvironment. *Cancer Cell Int* 2020; **20**: 106 [PMID: [32265595](#) DOI: [10.1186/s12935-020-01183-1](#)]
- 17 **Chen YJ**, Guo YN, Shi K, Huang HM, Huang SP, Xu WQ, Li ZY, Wei KL, Gan TQ, Chen G. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: a comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. *Respir Res* 2019; **20**: 48 [PMID: [30832674](#) DOI: [10.1186/s12931-019-0994-1](#)]
- 18 **Zhang Y**, Qiao WB, Shan L. Expression and functional characterization of FOXM1 in non-small cell lung cancer. *Oncotargets Ther* 2018; **11**: 3385-3393 [PMID: [29928129](#) DOI: [10.2147/OTT.S162523](#)]
- 19 **Wang L**, Wang Y, Du X, Yao Y, Wang L, Jia Y. MiR-216b suppresses cell proliferation, migration, invasion, and epithelial-mesenchymal transition by regulating FOXM1 expression in human non-small cell lung cancer. *Oncotargets Ther* 2019; **12**: 2999-3009 [PMID: [31114243](#) DOI: [10.2147/OTT.S202523](#)]
- 20 **Choi YE**, Madhi H, Kim H, Lee JS, Kim MH, Kim YN, Goh SH. FAM188B Expression Is Critical for Cell Growth via FOXM1 Regulation in Lung Cancer. *Biomedicines* 2020; **8** [PMID: [33142744](#) DOI: [10.3390/biomedicines8110465](#)]
- 21 **Xu C**, Li S, Chen T, Hu H, Ding C, Xu Z, Chen J, Liu Z, Lei Z, Zhang HT, Li C, Zhao J. miR-296-5p suppresses cell viability by directly targeting PLK1 in non-small cell lung cancer. *Oncol Rep* 2016; **35**: 497-503 [PMID: [26549165](#) DOI: [10.3892/or.2015.4392](#)]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-3991568

E-mail: bpgoffice@wjgnet.com

Help Desk: <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

