



OPEN ACCESS

EDITED BY

Ming Jun Zheng,
Ludwig Maximilian University of Munich,
Germany

REVIEWED BY

Feng Jiang,
Fudan University, China
Bo Wang,
First Affiliated Hospital of Chinese PLA
General Hospital, China
Yuyi Hou,
First Affiliated Hospital of Jiamusi
University, China

*CORRESPONDENCE

Shan-Shan Wang,
✉ swansea33@163.com
Jin-Hong Mei,
✉ mjhdoctor@126.com

SPECIALTY SECTION

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 08 January 2023

ACCEPTED 08 March 2023

PUBLISHED 16 March 2023

CITATION

Lai L-T, Ren Y-H, Huai Y-J, Liu Y, Liu Y,
Wang S-S and Mei J-H (2023),
Identification and validation of novel
prognostic biomarkers and therapeutic
targets for non-small cell lung cancer.
Front. Genet. 14:1139994.
doi: 10.3389/fgene.2023.1139994

COPYRIGHT

© 2023 Lai, Ren, Huai, Liu, Liu, Wang and
Mei. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identification and validation of novel prognostic biomarkers and therapeutic targets for non-small cell lung cancer

Li-Ting Lai^{1,2}, Yuan-Hui Ren^{2,3}, Ya-Jun Huai^{1,2}, Yu Liu^{2,3},
Ying Liu^{2,3}, Shan-Shan Wang^{2,3*} and Jin-Hong Mei^{2,3*}

¹Department of Oncology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China,

²Department of Pathology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China,

³Institute of Molecular Pathology, Nanchang University, Nanchang, Jiangxi, China

Background: Despite the significant survival benefits of anti-PD-1/PD-L1 immunotherapy, non-small cell lung cancer (NSCLC) remains one of the most common tumors and major causes of cancer-related deaths worldwide. Thus, there is an urgent need to identify new therapeutic targets for this refractory disease.

Methods: In this study, microarray datasets GSE27262, GSE75037, GSE102287, and GSE21933 were integrated by Venn diagram. We performed functional clustering and pathway enrichment analyses using R. Through the STRING database and Cytoscape, we conducted protein-protein interaction (PPI) network analysis and identified the key genes, which were verified by the GEPIA2 and UALCAN portal. Validation of actin-binding protein anillin (ANLN) was performed by quantitative real-time polymerase chain reaction and Western blotting. Additionally, Kaplan-Meier methods were used to compute the survival analyses.

Results: In total, 126 differentially expressed genes were identified, which were enriched in mitotic nuclear division, mitotic cell cycle G2/M transition, vasculogenesis, spindle, and peroxisome proliferator-activated receptor signaling pathway. 12 central node genes were identified in the PPI network complex. The survival analysis revealed that high transcriptional levels were associated with inferior survival in NSCLC patients. The clinical implication of ANLN was further explored; its protein expression showed a gradually increasing trend from grade I to III.

Conclusion: These Key genes may be involved in the carcinogenesis and progression of NSCLC, which may serve as useful targets for NSCLC diagnosis and treatment.

KEYWORDS

non-small cell lung cancer, gene expression omnibus, differentially expressed genes, protein-protein interaction, biological process

Introduction

Lung cancer is the most common cause of cancer-related death worldwide, wherein NSCLC accounts for 85% of lung cancer cases (Sung et al., 2021). An increased understanding of the biology and pathogenic genomic changes in NSCLC has led to advances and developments in its treatment. Particularly, the emergence of molecularly targeted therapies and immunotherapy has fundamentally changed the way NSCLC patients are treated (Jordan et al., 2017). A large number of genes have been recognized as drug targets and their molecular alterations, including epidermal growth factor receptor mutations, proto-oncogene receptor tyrosine kinase 1 rearrangements, anaplastic lymphoma kinase rearrangements, and BRAF V600E mutations, could predict the response to treatment (Stella et al., 2013). Testing for these genes is becoming increasingly routine and has yielded motivating results.

However, the incidence of rearrangement, fusion, or over-expression of these genes in NSCLC patients are very low, leading to limited availability of molecular targeted therapies for these genes. For example, aberrantly activations of ALK was found in approximately 4% of NSCLC tumors, and chromosomal rearrangement of ROS1 has been identified in approximately 1% of NSCLC patients (Wong et al., 2009; Gainor and Shaw, 2013). EGFR somatic activating mutations were found in approximately 20% of advanced NSCLC patients, and represented a paradigm for the use of tyrosine kinase inhibitors for subsets of cancer treatment. However, acquired resistance inevitably occurs in these cases (Yu et al., 2015). In addition, there is currently a very limited number of drug targets for other subtypes of lung cancer, such as squamous cell and large cell carcinoma, other than adenocarcinoma. Furthermore, targeted drugs developed for lung adenocarcinoma are basically ineffective for lung squamous cell carcinoma (Rekhtman et al., 2012). As for immunotherapy, the improvement in survival of lung cancer patients by blocking the immune checkpoint PD-1/PD-L1 is encouraging. However, only about 20% of patients benefit, and resistance is likely to develop after the initial response (Topalian et al., 2019). Thus, identifying potential gene targets or pathway alterations in this refractory disease is urgently needed.

Currently, the availability of information about the human genome and proteome, especially those that assist in the development of new anti-cancer agents, is largely dependent on advances in bioinformatics. As an enabling technology, bioinformatics bridges the gap between sequence information and clinical practice, and it has evolved into multiple ways to enable us not only to identify “driver” and “passenger” genes toward neoplasia, but also to comprehend genetic alterations and mechanisms in cancer (Mount and Pandey, 2005).

In this study, four microarrays, namely GSE27262, GSE75037, GSE102287, and GSE21933, were integrated and analyzed. Differentially expressed genes (DEGs) between NSCLC samples and corresponding normal specimens were analyzed. Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the DEGs were developed. The PPI network was developed by the Search Tool for the Retrieval of Interacting Genes (STRING) database. We screened out the key genes with the supreme connectivity in the network and evaluated their prognostic value, which would be helpful for further development of prognostic biomarkers and novel therapeutic targets for NSCLC patients.

TABLE 1 The composition of four different gene expression omnibus datasets.

GEO series	NSCLC	Normal	Total number
GSE75037	83	83	166
GSE21933	21	21	42
GSE27262	25	25	50
GSE102287	32	34	66

Abbreviations: NSCLC, non-small cell lung cancer; GEO, gene expression omnibus.

Materials and methods

Microarray datasets information

The National Center for Biotechnology Information Gene Expression Omnibus (GEO) is an open-access database for data regarding next-generation sequencing, microarray, and other forms of high-throughput gene data (Barrett et al., 2013), from which the microarray datasets of lung cancer samples and adjacent non-malignant samples (GSE27262, GSE75037, GSE102287, and GSE21933) were downloaded. Gene expression profiles of GSE27262 and GSE102287 were based on platform GPL570 [HG-U133_Plus_2] Affymetrix Human Genome Array, with 25 lung adenocarcinoma tissues *versus* 25 adjacent normal specimens and 32 NSCLC samples *versus* 34 normal samples, respectively. GSE75037 was based on platform GPL6884 HumanWG-6 v3.0 expression beadchip, including 83 adenocarcinomas and 83 adjacent normal samples. GSE21933 was based on platform GPL6254 Phalanx Human OneArray, including 21 NSCLC tissues and 21 matched adjacent non-malignant tissues.

Data analysis

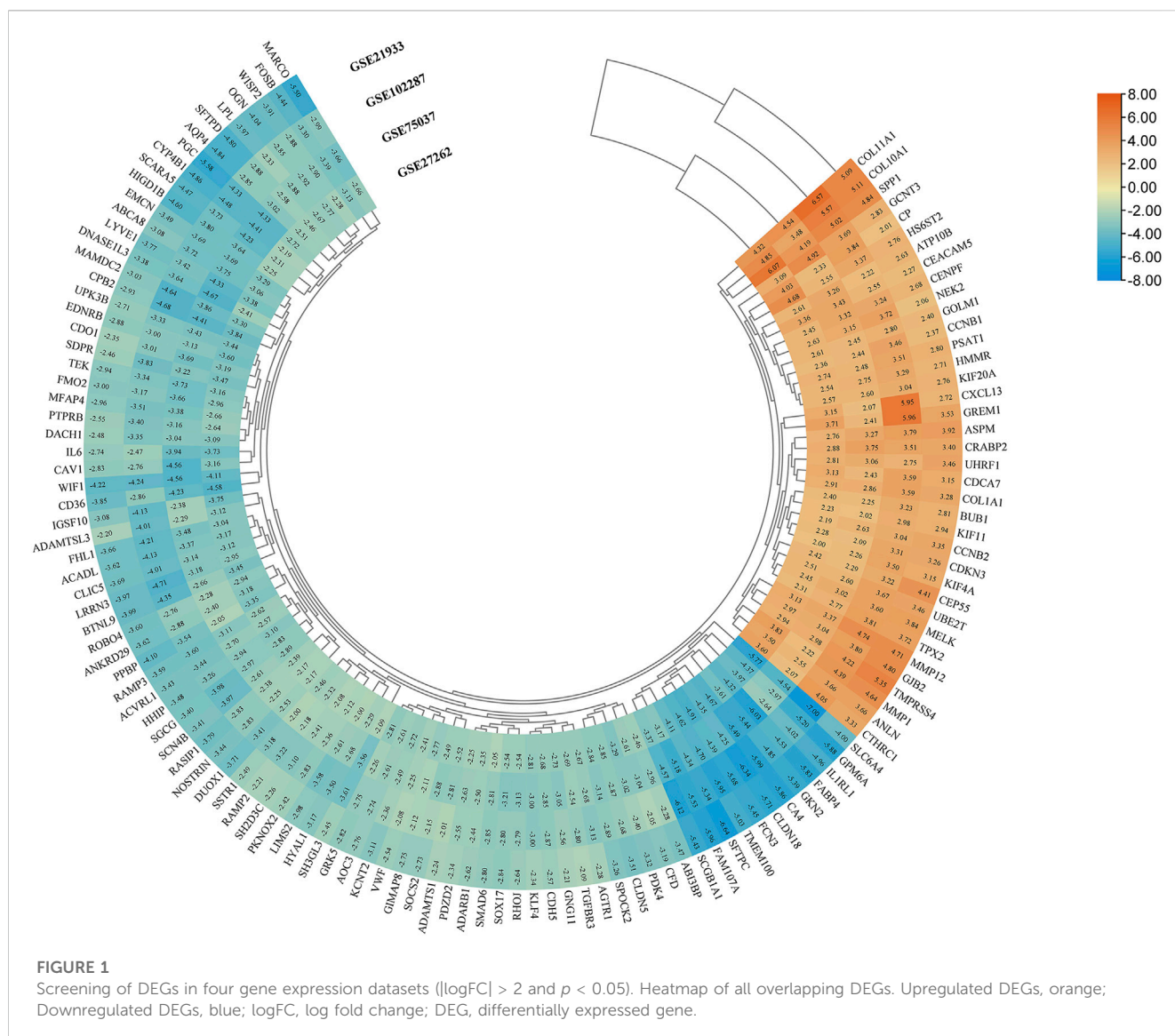
GEO2R, a network application based on R that utilizes the Bioconductor (R packages) to analyze GEO data, was used to identify DEGs between lung cancer and adjacent non-malignant specimens. The selection criteria, $|\log_{2}FC| > 2.0$, and adjusted $p < 0.05$ were used to define the DEGs. We analyzed each dataset and intersected them using Venn diagrams.

GO and KEGG pathway enrichment analysis

The GO knowledgebase was composed of ontology and ontology annotations. As of 2018, there were approximately 45,000 terms in GO, including CC, BP, and MF terms (The Gene Ontology Consortium, 2017). R software version 4.0.3 (clusterProfiler and ggplot2 packages) was used for gene classification and GO, KEGG pathway enrichment analyses. Statistical significance was set at $p < 0.05$.

PPI network visualization

STRING v11, an online resource with currently the largest number of proteins (24.6 million) and broad data sources



(Szklarczyk et al., 2019), was employed to explore protein-protein associations among the DEGs. In addition, Cytoscape software was used for the visualization of the protein interaction network and the analysis of the interaction of the candidate DEGs that encode proteins in NSCLC. The top 12 molecules with the strongest connectivity in the network were identified as key genes by CytoHubba, a plug-in of Cytoscape.

Genetic alteration analysis and enrichment analysis of the key gene-related drugs

Through the data of lung adenocarcinoma and lung squamous cell carcinoma of the TCGA project and the Sangerbox platform, we obtained the mutation profile of 12 key genes in NSCLC. Furthermore, we have enriched and analyzed these key gene-related drugs by using Enrichr platforms (<https://maayanlab.cloud/Enrichr/>).

We used Diseases/Drugs and DSigDB module for cluster analysis.

Survival analysis

To evaluate the effect of the 12 key genes on prognosis of NSCLC patients, we used the Kaplan–Meier plotter (<http://kmplot.com/analysis/>), an interactive database for validation of prognostic biomarkers that contains mRNA, miRNA, protein data, and clinical information from a variety of cancer patients. Patients with NSCLC were grouped based on their mRNA levels and hazard ratios, and the respective p values were calculated. In addition, we verified the survival analysis using the TCGA database (<https://portal.gdc.cancer.gov/>). We downloaded and collated lung adenocarcinoma and squamous cell carcinoma RNAseq data and clinical data from the TCGA database; Survival package of R software was used to test the

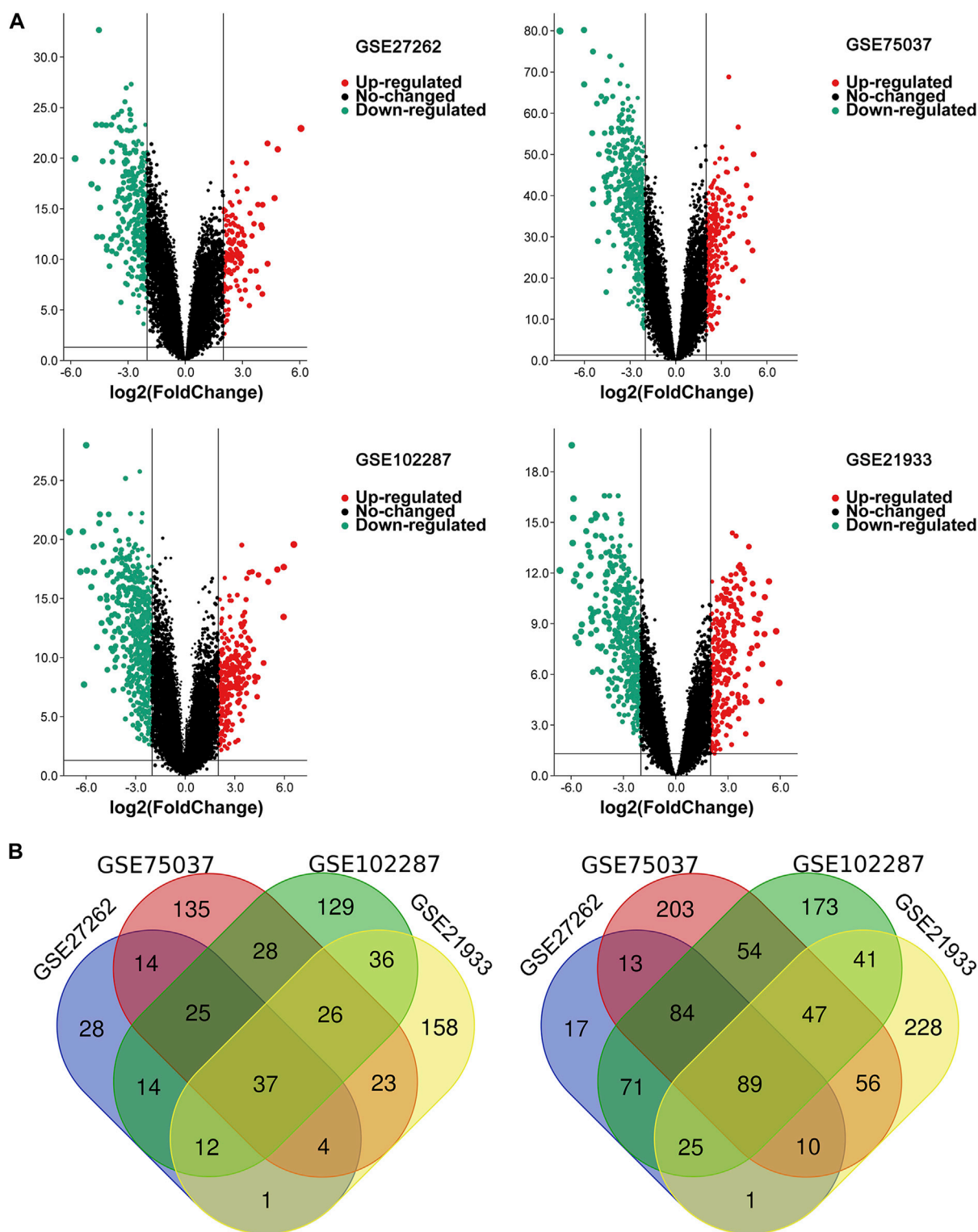
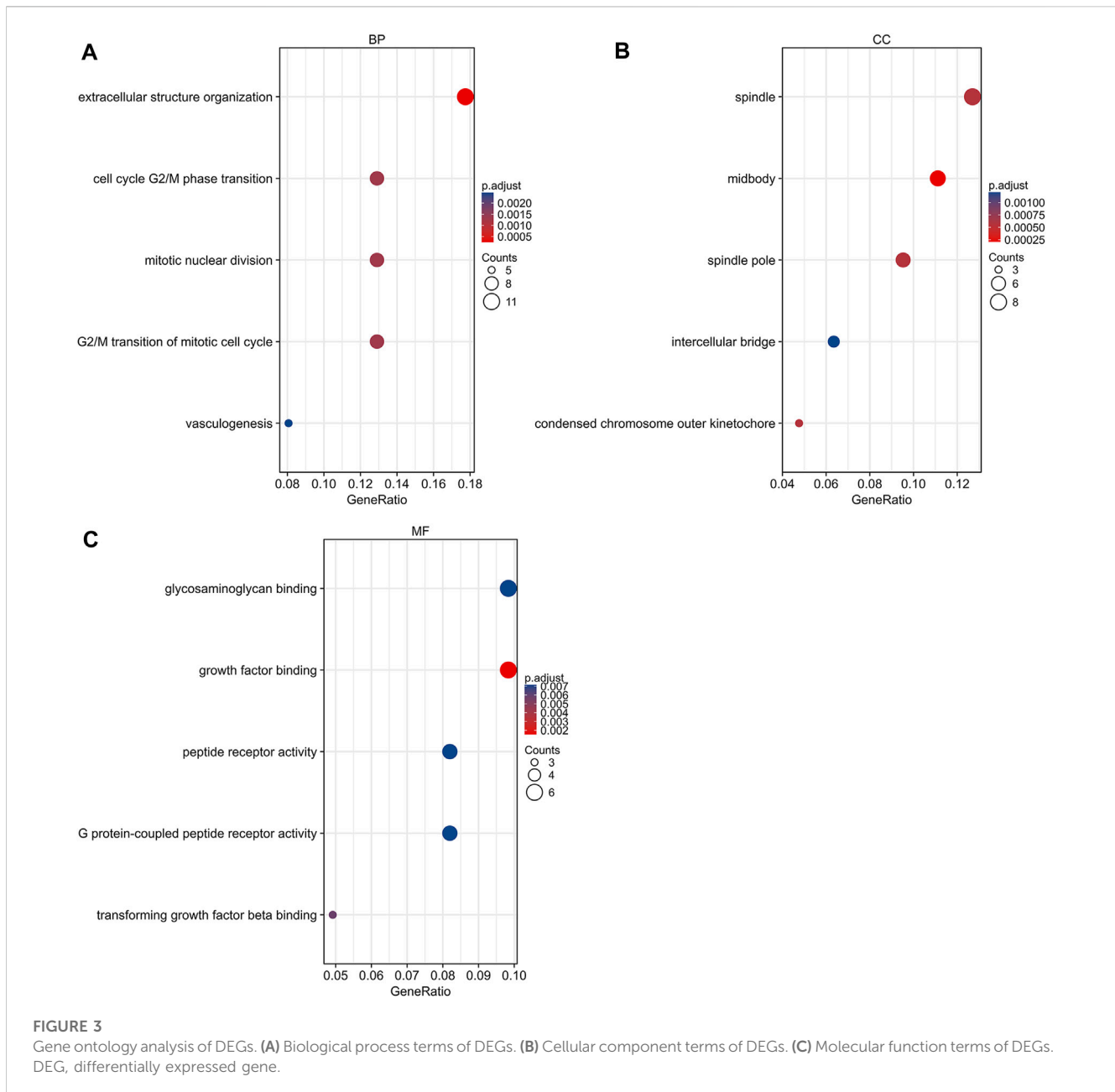


FIGURE 2
 The overlapping DEGs of the four gene expression datasets. **(A)** Volcano plots of each gene expression profiles in NSCLC and normal tissues. **(B)** Venn diagrams of DEGs. The one on the left refers to 37 upregulated DEGs; The right one refers to 89 downregulated DEGs; NSCLC, non-small cell lung cancer; DEG, differentially expressed gene.



proportional risk hypothesis, and the results were visualized using survminer package and ggplot2 package.

Expression analysis and clinicopathological association

The expression validation of the key genes was performed based on RNA sequencing data produced by The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) project. Tissue-wise expression analyses of key genes between 969 NSCLC samples and 685 non-malignant samples from TCGA and the GTEx project were profiled using GEPIA2 (<http://gepia2.cancer-pku.cn/>).

Clinicopathologic features of patients with NSCLC, including pathologic stage, tumor grade, age, gender, living status, and body

weight, were obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Confirmatory/Discovery dataset. Proteomic analyses of lung cancer and normal samples were performed using UALCAN, an open network repository for investigation on gene expression and its disease association (Chandrashekar et al., 2017). Furthermore, we obtained its immunohistochemical results from the HPA database (The Human Protein Atlas <https://www.proteinatlas.org/>).

Cell culture

Lung cancer cell lines NCI-H1975, NCI-H1650, A549, and NCI-H1299, and normal lung epithelial cell line BEAS-2B were bought from the Shanghai Cell Bank and ICCELL Company. Cells were

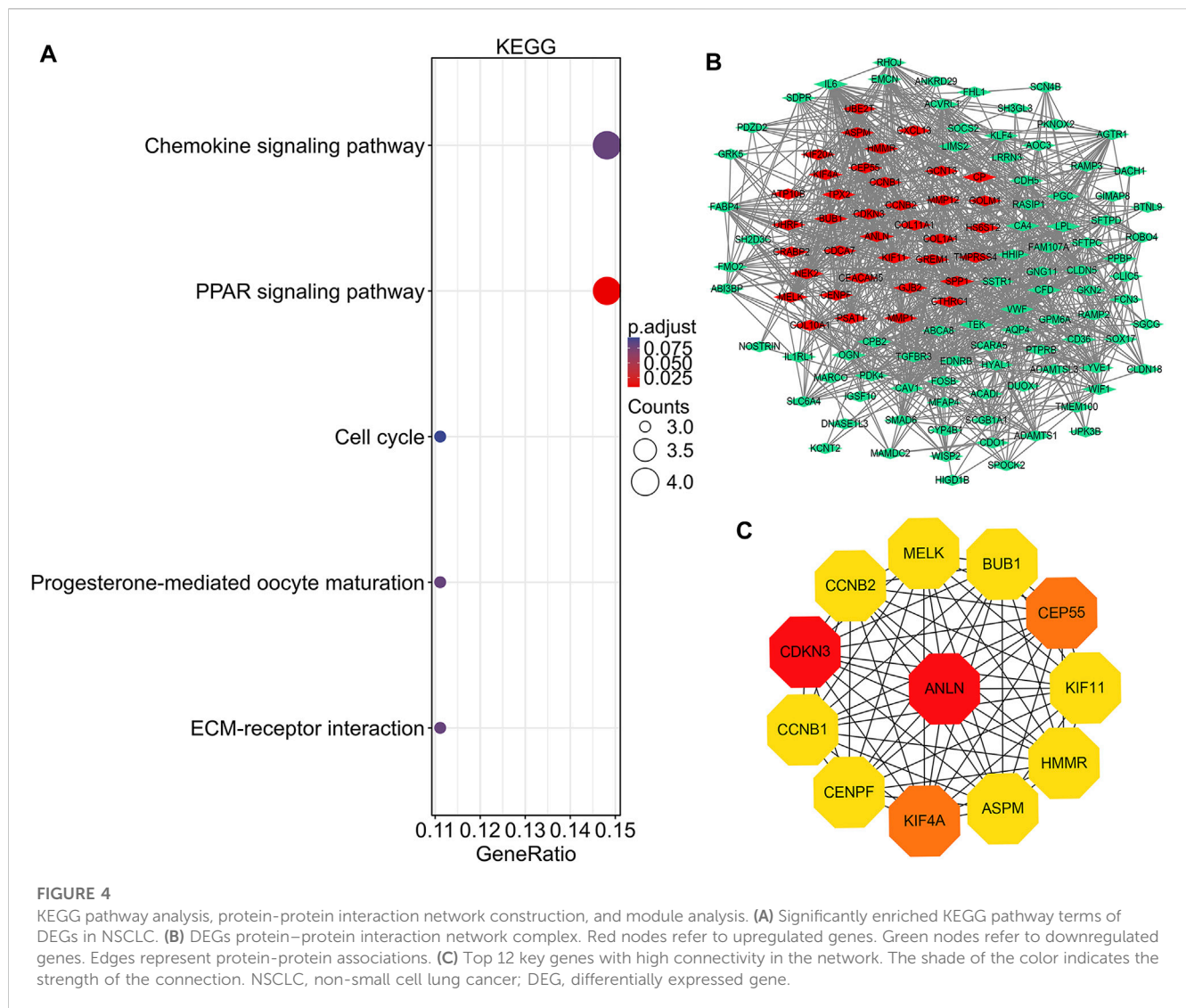


FIGURE 4

KEGG pathway analysis, protein-protein interaction network construction, and module analysis. **(A)** Significantly enriched KEGG pathway terms of DEGs in NSCLC. **(B)** DEGs protein-protein interaction network complex. Red nodes refer to upregulated genes. Green nodes refer to downregulated genes. Edges represent protein-protein associations. **(C)** Top 12 key genes with high connectivity in the network. The shade of the color indicates the strength of the connection. NSCLC, non-small cell lung cancer; DEG, differentially expressed gene.

cultured in Roswell Park Memorial Institute-1640 (RPMI-1640; Solarbio, Beijing, China) and Dulbecco's modified Eagle's medium (DMEM; Solarbio, Beijing, China) supplemented with 10% fetal bovine serum (FBS; Gemini, California, United States) and maintained at 37 °C thermostatic and humidified cell incubator with 5% CO₂.

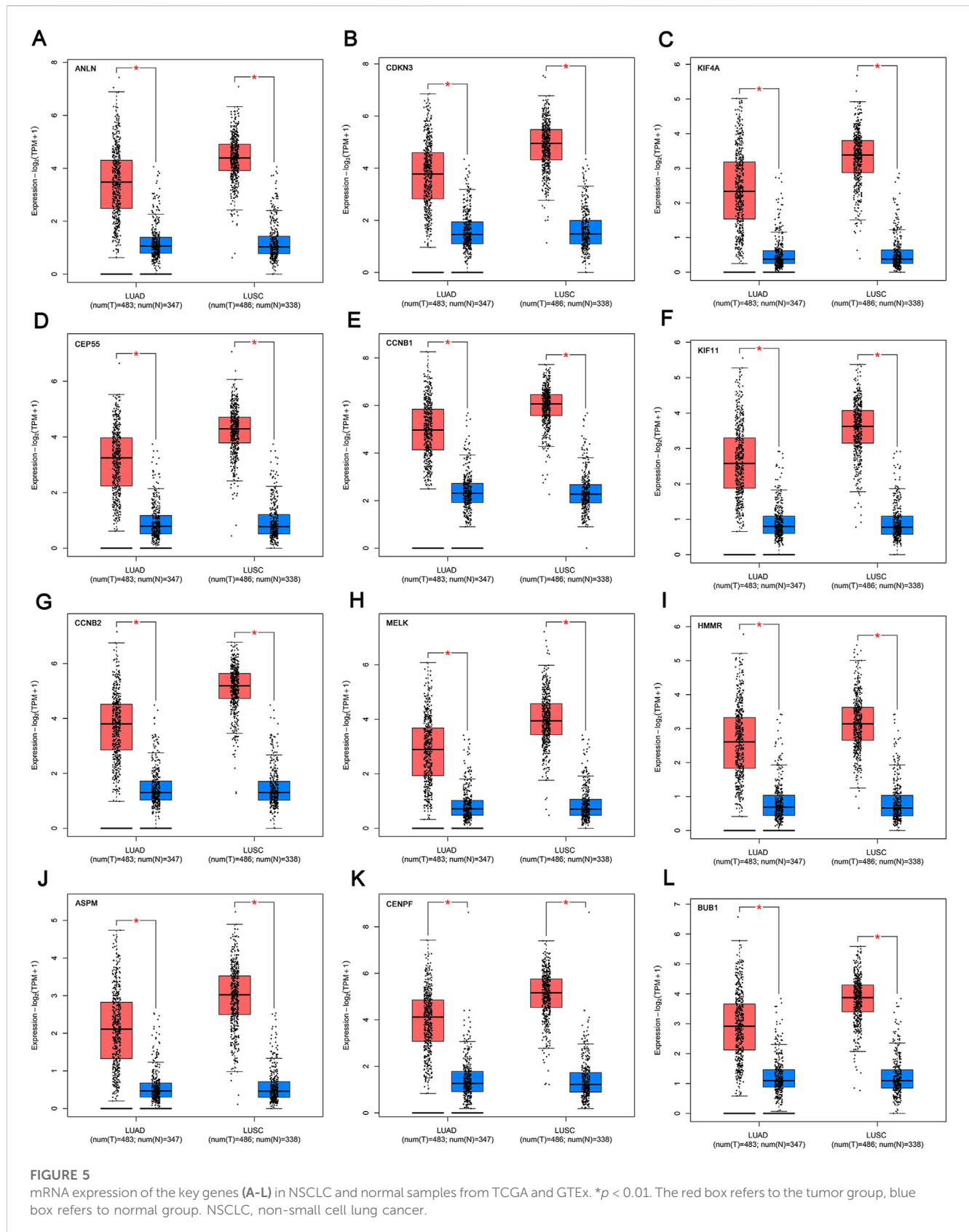
RNA extraction and qRT-PCR

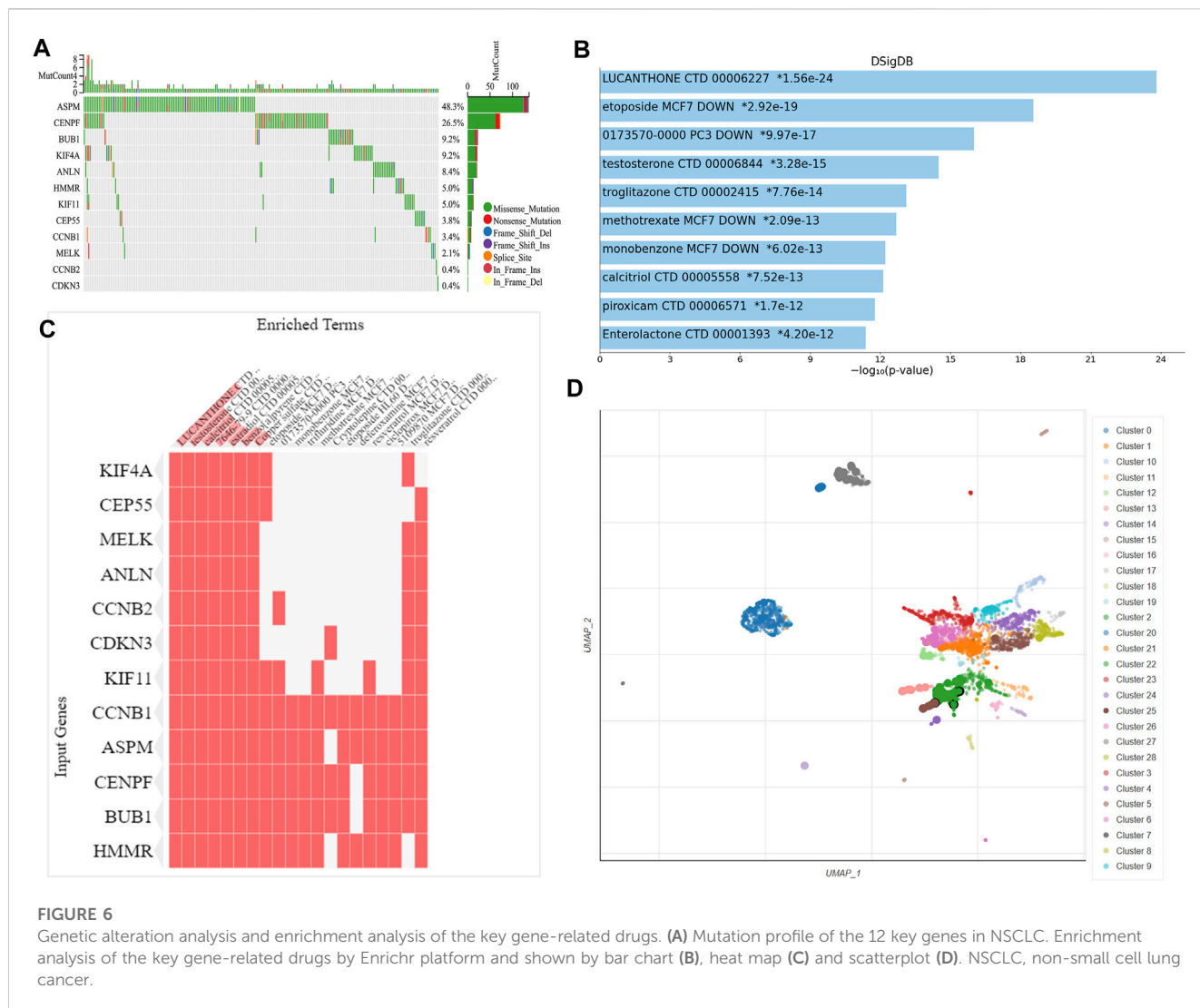
Total RNA was extracted from NSCLC cell lines and normal lung epithelial cell lines with an RNA extraction kit (Axygen; Silicon Valley, United States) and reverse transcription was performed using a cDNA Synthesis Kit (Vazyme Biotech, Nanjing, China). Quantitative real-time polymerase chain reaction (qRT-PCR) was carried out using Bio-Rad CFX96 Touch with ChamQ SYBR[®] Green qRT-PCR Master Mix. All qRT-PCRs were performed three times and measured using $2^{-\Delta\Delta CT}$ algorithm. Primer sequences were as follows: ANLN, Former: TCTTCGTGGCCGATTTGACA, Reverse: TGGACTTACCACCAACTGT; GAPDH, Former: CGAGCC

ACATCGCTCAGACA, Reverse: GTGGTGAAGACGCCA GTGGA.

Western blotting

We extracted proteins for Western blotting using RIPA lysis buffer (Solarbio, Beijing, China; R0010) and phenylmethylsulfonyl fluoride protease inhibitors (Solarbio, Beijing, China; IP0280). The BCA Protein Assay Kit (Vazyme Biotech, Nanjing, China; E112-02) was used for protein concentration determination. The Western blot system was established using the Bio-Rad Bis-Tris Gel system according to the manufacturer's instructions. Proteins isolated by SDS-PAGE were electroblotted onto polyvinylidene fluoride membranes and incubated with a primary antibody (dilution: 1:1250) overnight in a shaker at 4°C. They were then incubated in a shaker for 1 h with horseradish peroxidase labeled secondary antibody (dilution: 1:20000) at 25°C. After rinsing, a multi-functional chemiluminescent imaging system (Analytik-Jena, United States) was used for development.





Results

Screening of DEGs

Four microarray datasets (GSE27262, GSE75037, GSE102287, and GSE21933) were selected in this study, and their statistics are shown in Table 1. Clustering of all overlapping DEGs is shown in the heatmap (Figure 1). In accordance with the selection criteria, $|\log_2FC| > 2.0$ and adjusted $p < 0.05$, a total of 445, 845, 891, and 794 DEGs were identified from the GSE27262, GSE75037, GSE102287, and GSE21933 microarrays, respectively, as shown in the volcano plots (Figure 2A). After intersecting the DEGs of the four databases, 126 DEGs including 37 upregulated genes and 89 downregulated genes, were found to be significant in all four microarray datasets (Figure 2B).

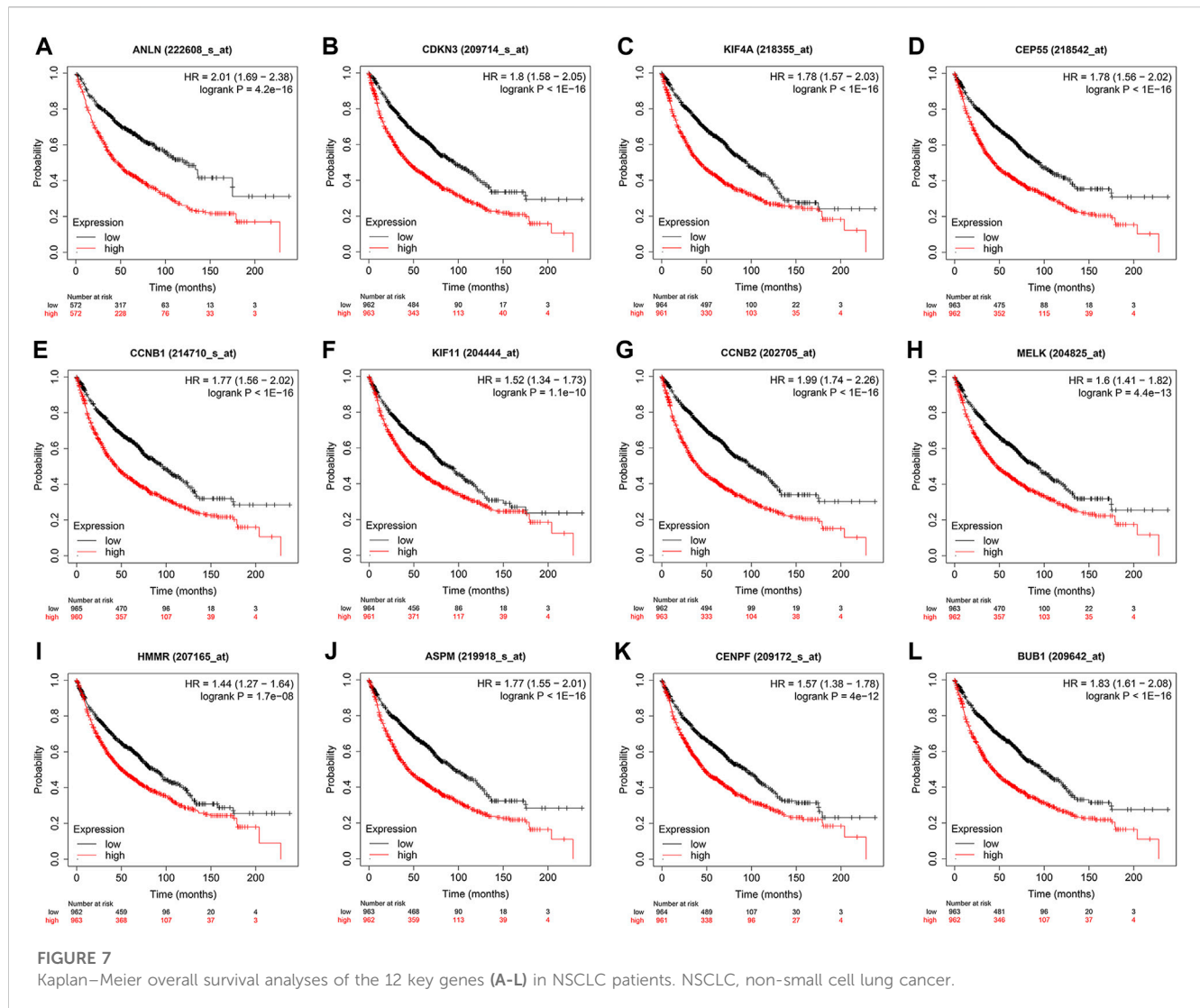
Functional annotation and pathway enrichment analyses

GO and KEGG pathway analyses were conducted using R 4.0.3 (clusterProfiler, org.Hs.e.g.,db, and ggplot2 packages). DEGs were

basically enriched in mitotic nuclear division, cell cycle, G2/M phase transition, vasculogenesis, G2/M transition of mitotic cell cycle in biological process (BP) terms, spindle, midbody, condensed chromosome outer kinetochore in cellular components (CC) terms, and growth factor binding, G protein-coupled peptide receptor activity, and peptide receptor activity in molecular functions (MF) terms (Figure 3). The KEGG pathway analysis found that the DEGs were predominantly involved in the peroxisome proliferator-activated receptors (PPAR) signaling pathway, cell cycle, and ECM-receptor interaction pathway (Figure 4A).

PPI network construction and key gene identification

There were 126 nodes and 1,054 edges in the PPI network with an enrichment p -value of $< 1.0e-16$ (Figure 4B). Twelve central node genes, including ANLN, cyclin-dependent kinase inhibitor 3 (CDKN3), kinesin family member 4A (KIF4A), centrosomal protein 55 kDa (CEP55), G2/mitotic-specific cyclin-B1



(CCNB1), kinesin family member 11 (KIF11), G2/mitotic-specific cyclin-B2 (CCNB2), maternal embryonic leucine zipper kinase (MELK), hyaluronan-mediated motility receptor (HMMR), abnormal spindle-like microcephaly associated protein (ASPM), centromere protein F (CENPF), and checkpoint serine/threonine-protein kinase (BUB1), were identified among the 126 nodes by using CytoHubba of Cytoscape (Figure 4C). Furthermore, ANLN was the top gene in the network with the highest connectivity and maximum neighborhood component (Table S1).

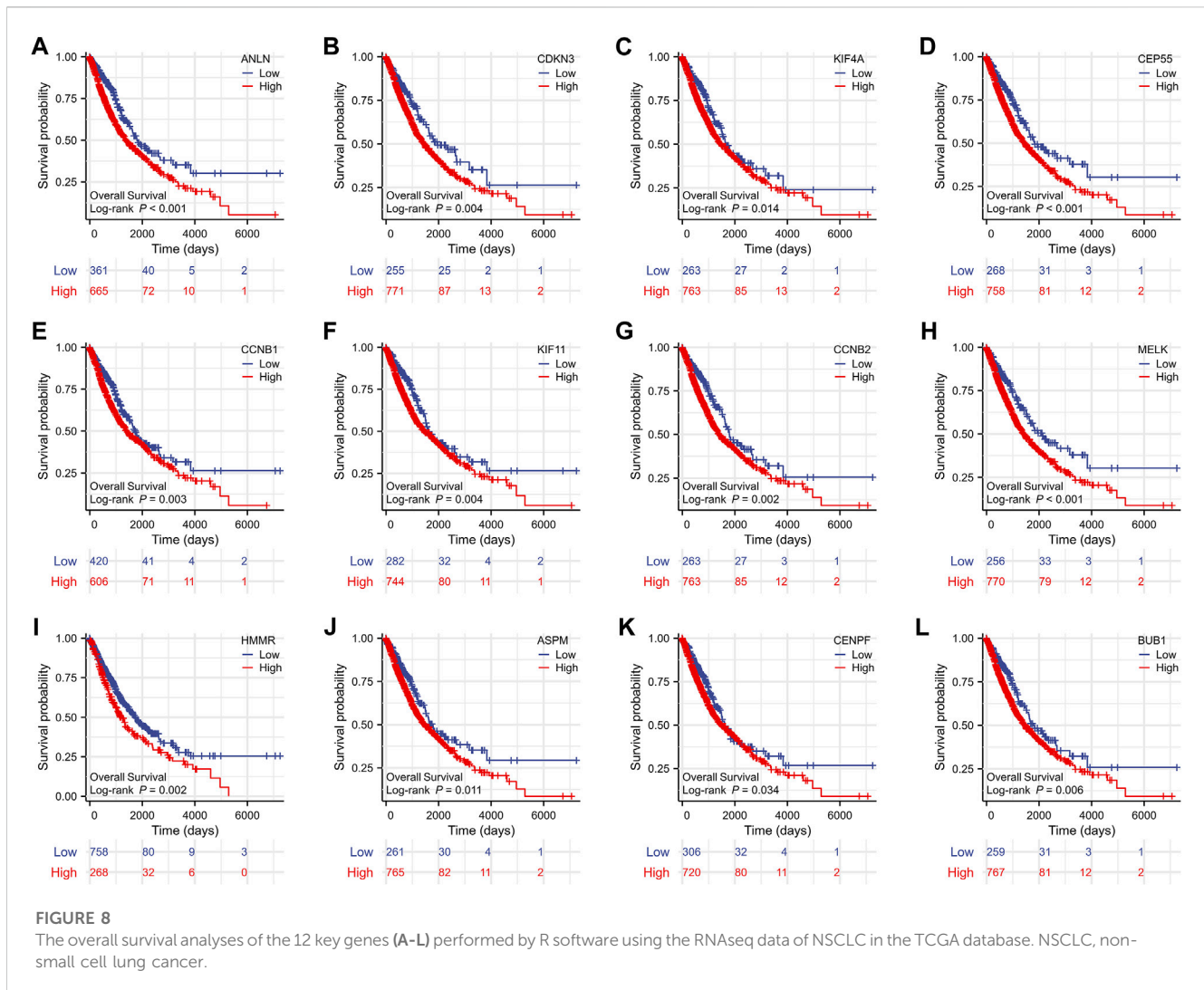
Transcriptional level validation of the 12 key genes

We profiled the tissue-wise expression of key genes in NSCLC tissues and normal tissues using GEPIA2. The results revealed that the mRNA expression levels of the 12 key genes in the NSCLC

samples were significantly higher than those in normal samples (Figure 5).

Genetic alteration analysis and key gene-related drugs enrichment analysis

We observed the mutation status of these key genes in different NSCLC samples of TCGA. As shown in Figure 6A, ASPM had the highest mutation frequency, followed by CENPF. CCNB2 and CDKN3 had the lowest mutation frequency. Missense mutation and frame-shift mutation were the most common types of mutations, while in-frame internal deletion was rare. To explore drugs that associated with the key genes, we used Enrichr platform to perform cluster analysis and UMAP algorithm to draw scatter map for all corresponding terms in DSigDB gene set database. We found that the terms of enrichment of these key genes were correlated with antitumor drugs etoposide and



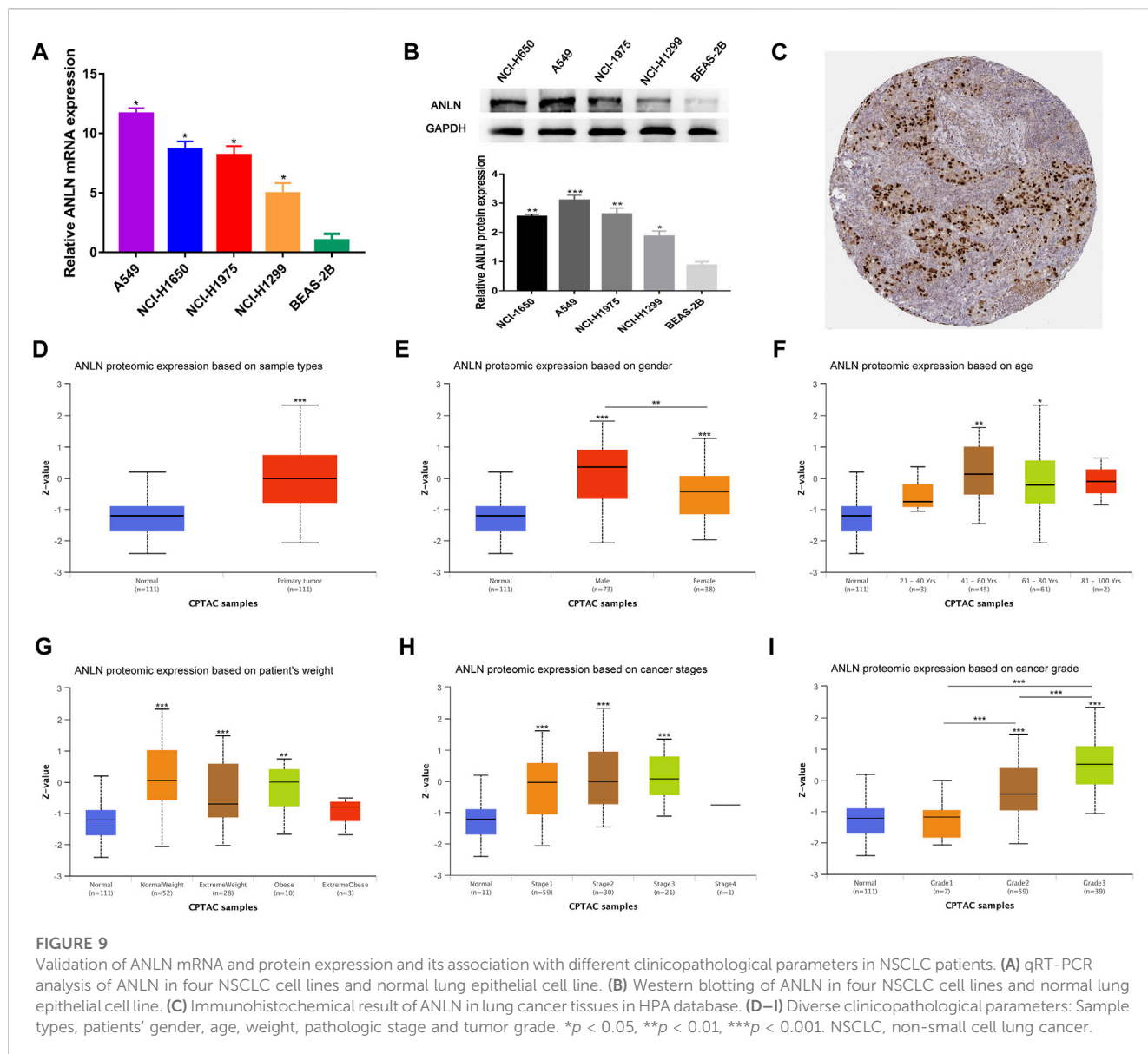
methotrexate, as well as non-tumor drugs such as lucanthone, troglitazone, testosterone, calcitriol and piroxicam (Figure 6).

Prognostic role of key genes

To evaluate the prognostic values of the 12 key genes, we used the Kaplan–Meier plotter, an online database that contained transcriptomic data of 3,452 NSCLC patients. Just as PD-L1 expression, tumor mutational burden can be used to predict immune checkpoint inhibitor outcomes, the key molecules we identified can predict survival outcomes in patients with NSCLC. Overall survival (OS) and first-progression (FP) survival curves are shown in Figure 7; Supplementary Figure S1. High transcriptional levels of the 12 key genes (ANLN, CDKN3, KIF4A, CEP55, CCNB1, KIF11, CCNB2, MELK, HMMR, ASPM, CENPF, and BUB1) were all significantly related to poorer OS (all $p < 0.001$) and FP survival (all $p < 0.01$) in NSCLC. We verified the overall survival analysis of NSCLC patients through the TCGA database, and the conclusion reached was consistent with those obtained by the Kaplan–Meier plotter analysis (Figure 8).

In vitro verification of ANLN and the relationship between its protein expression and the clinicopathologic parameters of NSCLC patients

To verify the transcription level and protein expression level of ANLN, QRT-PCR and Western blotting assays were performed in BEAS-2B and four NSCLC cell lines. We found that both mRNA and protein levels of ANLN in the four NSCLC cell lines were significantly higher than those in BEAS-2B (Figures 9A, B). Through the HPA database, we found that ANLN was strongly positive in the immunohistochemical test of lung cancer tissues (Figure 9C). In addition, we further investigated the relationship of ANLN and various clinicopathological parameters of NSCLC and its gene expression profile in different cancer types. There was a gradually increasing trend based on the protein expression of ANLN from grade I to grade III, while age, weight, and tumor stage groups did not significantly differ given the protein expression of ANLN (Figures 9D–I). Interestingly, we also found that ANLN level of was higher in male patients than in female patients (Figure 9E, $p < 0.01$). As shown in Supplementary Figure S1,



ANLN is elevated in various TCGA and GTEx tumors, including hepatocellular carcinoma, pancreatic adenocarcinoma, and breast carcinoma, compared with paired normal tissues.

Discussion

With rapidly increasing morbidity and mortality, the 5-year survival of lung cancer patients varies from 4% to 17%, depending on the region and stage (Hirsch et al., 2017). Substantial progress has been made in NSCLC treatment in recent years, but long-term effective responses are still rare for most patients (Herbst et al., 2018). It remains critical to explore the underlying pathogenesis of lung cancer and achieve more precise treatment. A number of researchers have made impressive progress in this area, exploring the microenvironment of tumors, looking for biomarkers and individual targeted treatment strategies (Guo et al., 2022; Jiang et al., 2022).

Rather than focusing on a single cohort study, we integrated four cohorts of microarray databases and identified 126 overlapping DEGs (37 upregulated and 89 downregulated) in this study. Through further functional clustering and enrichment analyses, we found that these genes were mainly enriched in the mitotic nuclear division, cell cycle G2/M phase transition, and PPAR signaling pathway. Mitotic nuclear division, a biological process that is complementary to but opposite to apoptosis, plays a crucial part in carcinogenesis, tumor cell maintenance, and tumor progression (Sinha et al., 2019). Given that cancer is a cell cycle disease, the progression of the cell cycle is inextricably linked to the proliferation and activation of cancer cells. The progression of the cell cycle is coordinated by the continuous activation of cyclin-dependent kinases through their corresponding cyclin chaperone (Malumbres, 2014). Some tumor suppressor genes and drug molecules can inhibit tumor cell proliferation and invasion by arresting the cell in the G2/M phase transition (Song et al., 2009). PPARs have three subtypes (PPAR- α , PPAR- β and PPAR- γ), which exhibit diverse roles in vertebrates. PPAR- α mainly plays a role in

removing circulating lipids or cell lipids, PPAR- β is involved in lipid oxidation and cell proliferation, while PPAR- γ activation enhances the proliferation of cancer cells and promotes brain metastasis (Bougarne et al., 2018; Magadum and Engel, 2018; Zou et al., 2019). To further explore the internal interactions of the overlapping DEGs, a PPI network was developed. 12 genes with the strongest connectivity in the network were identified. High transcriptional levels of these genes were significantly correlated with poor prognosis, which reveals their potential prognostic value.

ANLN, the top gene in our modules, is a unique scaffolding protein that was first isolated from *Drosophila melanogaster* embryos and was mainly associated with cytokinesis (Zhang and Maddox, 2010). ANLN has been reported to be overexpressed in many tumors. It is involved in the progression of pancreatic, brain, breast, and lung cancers (Hall et al., 2005; Olakowski et al., 2009; Magnusson et al., 2016; Long et al., 2018), which is consistent with our experimental results. Evidence has shown that ANLN promotes cell proliferation, and the loss of ANLN prevents the cancer cells from dividing and reduces their migration and invasion (Wang et al., 2019). Furthermore, there is also evidence showing that ANLN expression correlates with lung adenocarcinoma metastasis (Xu et al., 2019). In breast cancer, ANLN was found to be an alternative marker for Ki-67 (cell proliferation index), which is consistent with our findings (Figure 6F). Based on the evidence supporting the correlation of ANLN with acknowledged features of cancer, ANLN should be considered as a novel target for cancer therapy.

CDKN3 has been reported to be overexpressed in glioma and cervical cancer, and its over-expression is associated with inferior survival (Yu et al., 2007; Espinosa et al., 2013). Since there are more mitotic cells in rapidly proliferating tumor cells, CDKN3 transcription and protein levels fluctuate throughout the cell cycle, reaching a peak during mitosis. High levels of mitotic CDKN3 expression is the most likely mechanism for frequent CDKN3 mRNA over-expression in human cancer (Fan et al., 2015). The cell cycle-dependent elements of CCNB1 and CCNB2 are essential for meiotic resumption. CCNB1 has been observed to expedite tumor cell division, cell proliferation, and tumor growth in colorectal and pancreatic cancers (Fang et al., 2014; Zhang et al., 2018). CCNB2 is also correlated with cancer progression and inferior prognosis in breast cancer, hepatocellular carcinoma and NSCLC (Qian et al., 2015; Li et al., 2019; Jayanthi et al., 2020). KIF4A, the kinesin family member 4A, plays a key role in process of DNA replication and repair. It promotes cell proliferation, correlates with the size of the tumor in oral carcinoma, and serve as a potential prognostic indicator in various solid tumors (Wu et al., 2008; Rouam et al., 2010). KIF11 (E.g.,5) and MELK have been identified as oncogenes in multiple tumors and inhibiting agents targeting them have entered phase I/II clinical trials with encouraging results (Ganguly et al., 2014; Garcia-Saez and Skoufias, 2021). As of now, nine clinical trials targeting KIF11 have been completed, and five clinical trials targeting MELK are ongoing or completed, according to [ClinicalTrials.gov](https://clinicaltrials.gov/) (<https://clinicaltrials.gov/>). These drugs are used alone or in combination with other medicines to treat patients with refractory cancers.

CEP55 was identified as an ideal cancer vaccine candidate (Inoda et al., 2011) and a marker for predicting cancer invasion risk, metastasis, and therapeutic outcome (Tandon and Banerjee, 2020). HMMR, alternatively called RHAMM or CD168, is a

microtubule-associated protein that regulates mitosis and meiosis. (Chen et al., 2018). Abnormal expression of HMMR disrupts the microtubule process during cell division and leads to abnormalities in the mitotic spindle, altering the fate of progenitor cells and leading to genomic instability (Pujana et al., 2007). HMMR has been reported to be closely linked to cancer risk and progression in various tumor types (Rein et al., 2003). Currently, there are limited researches on ASPM's role in tumors. Recently, it has been reported as a new predictor of tumor aggressiveness and prognosis in bladder, prostate, and endometrial cancers. (Pai et al., 2019; Saleh et al., 2020; Zhou et al., 2020). The prenylated protein CENPF has been used clinically as a proliferative marker for malignant tumor cell growth (Varis et al., 2006). BUB1, a serine/threonine-protein kinase, plays a crucial part in oncogenesis, chromosome arrangement, and spindle assembly (Bolanos-Garcia and Blundell, 2011).

Finally, we profiled the tissue-specific expression of key genes in NSCLC and normal specimens from TCGA database and found that its mRNA levels were significantly elevated in tumor than in adjacent non-tumor tissues. We further explored the clinical implication of ANLN, and its protein expression showed a gradually increasing trend from grade I to III, revealing its association with tumor aggressiveness.

Conclusion

Through multiple microarray datasets and integrated bioinformatics analysis, we identified key genes and pathways that may be involved in NSCLC carcinogenesis, which are mainly associated with mitosis, vasculogenesis, and G2/M transition of the mitotic cell cycle. These findings provide new insights and opportunities for further development of prognostic biomarkers and therapeutic targets for NSCLC patients.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding authors.

Author contributions

J-HM and S-SW designed the study. Y-HR, Y-JH, YUL, and YIL collected the data. L-TL performed the experiments and wrote the article. Data analysis and interpretation was performed by L-TL and Y-JH. All authors read and approved the submitted version.

Funding

This work was funded by the National Natural Science Foundation of China (No. 81560410); Science and Technology Project of Jiangxi Provincial Administration of Traditional Chinese Medicine (No. 2020B0365); Science and Technology Planning Project of Jiangxi Provincial Health Department (No. 20203122).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1139994/full#supplementary-material>

References

- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomaszewski, M., et al. (2013). Ncbi geo: Archive for functional genomics data sets--update. *Nucleic acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Bolanos-Garcia, V. M., and Blundell, T. L. (2011). Bub1 and Bubr1: Multifaceted kinases of the cell cycle. *Trends Biochem. Sci.* 36 (3), 141–150. doi:10.1016/j.tibs.2010.08.004
- Bougarne, N., Weyers, B., Desmet, S. J., Deckers, J., Ray, D. W., Staels, B., et al. (2018). Molecular actions of ppara in lipid metabolism and inflammation. *Endocr. Rev.* 39 (5), 760–802. doi:10.1210/er.2018-00064
- Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). Ualcan: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia (New York, NY)* 19 (8), 649–658. doi:10.1016/j.neo.2017.05.002
- Chen, H., Connell, M., Mei, L., Reid, G. S. D., and Maxwell, C. A. (2018). The nonmotor adaptor hmmr dampens eg5-mediated forces to preserve the kinetics and integrity of chromosome segregation. *Mol. Biol. Cell* 29 (7), 786–796. doi:10.1091/mbc.E17-08-0531
- Espinosa, A. M., Alfaro, A., Roman-Basaure, E., Guardado-Estrada, M., Palma, L., Serralde, C., et al. (2013). Mitosis is a source of potential markers for screening and survival and therapeutic targets in cervical cancer. *PLoS one* 8 (2), e55975. doi:10.1371/journal.pone.0055975
- The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic acids Res.* 45(1), D331–D8. doi:10.1093/nar/gkw1108
- Fan, C., Chen, L., Huang, Q., Shen, T., Welsh, E. A., Teer, J. K., et al. (2015). Overexpression of major Cdkn3 transcripts is associated with poor survival in lung adenocarcinoma. *Br. J. cancer* 113 (12), 1735–1743. doi:10.1038/bjc.2015.378
- Fang, Y., Yu, H., Liang, X., Xu, J., and Cai, X. (2014). Chk1-Induced Ccnb1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. *Cancer Biol. Ther.* 15 (9), 1268–1279. doi:10.4161/cbt.29691
- Gainor, J. F., and Shaw, A. T. (2013). Novel targets in non-small cell lung cancer: Ros1 and ret fusions. *Oncol.* 18 (7), 865–875. doi:10.1634/theoncologist.2013-0095
- Ganguly, R., Hong, C. S., Smith, L. G., Kornblum, H. I., and Nakano, I. (2014). Maternal embryonic leucine zipper kinase: Key kinase for stem cell phenotype in glioma and other cancers. *Mol. cancer Ther.* 13 (6), 1393–1398. doi:10.1158/1535-7163.mct-13-0764
- Garcia-Saez, I., and Skoufias, D. A. (2021). Eg5 targeting agents: From new anti-mitotic based inhibitor discovery to cancer therapy and resistance. *Biochem. Pharmacol.* 184, 114364. doi:10.1016/j.bcp.2020.114364
- Guo, C. R., Mao, Y., Jiang, F., Juan, C. X., Zhou, G. P., and Li, N. (2022). Computational detection of a genome instability-derived lncrna signature for predicting the clinical outcome of lung adenocarcinoma. *Cancer Med.* 11 (3), 864–879. doi:10.1002/cam4.4471
- Hall, P. A., Todd, C. B., Hyland, P. L., McDade, S. S., Grabsch, H., Dattani, M., et al. (2005). The septin-binding protein anillin is overexpressed in diverse human tumors. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 11 (19), 6780–6786. doi:10.1158/1078-0432.ccr-05-0997
- Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553 (7689), 446–454. doi:10.1038/nature25183
- Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr., Wu, Y. L., et al. (2017). Lung cancer: Current therapies and new targeted treatments. *Lancet (London, Engl.)* 389 (10066), 299–311. doi:10.1016/s0140-6736(16)30958-8
- Inoda, S., Morita, R., Hirohashi, Y., Torigoe, T., Asanuma, H., Nakazawa, E., et al. (2011). The feasibility of Cep55/C10orf3 derived peptide vaccine therapy for colorectal carcinoma. *Exp. Mol. pathology* 90 (1), 55–60. doi:10.1016/j.yexmp.2010.10.001
- Jayanthi, V., Das, A. B., and Saxena, U. (2020). Grade-specific diagnostic and prognostic biomarkers in breast cancer. *Genomics* 112 (1), 388–396. doi:10.1016/j.ygeno.2019.03.001
- Jiang, F., Hu, Y., Liu, X., Wang, M., and Wu, C. (2022). Methylation pattern mediated by M(6)a regulator and tumor microenvironment invasion in lung adenocarcinoma. *Oxidative Med. Cell. Longev.* 2022, 2930310. doi:10.1155/2022/2930310
- Jordan, E. J., Kim, H. R., Arcila, M. E., Barron, D., Chakravarty, D., Gao, J., et al. (2017). Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discov.* 7 (6), 596–609. doi:10.1158/2159-8290.cd-16-1337
- Li, R., Jiang, X., Zhang, Y., Wang, S., Chen, X., Yu, X., et al. (2019). Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis. *Archives Med. Res.* 50 (1), 10–17. doi:10.1016/j.arcmed.2019.03.003
- Long, X., Zhou, W., Wang, Y., and Liu, S. (2018). Prognostic significance of anln in lung adenocarcinoma. *Oncol. Lett.* 16 (2), 1835–1840. doi:10.3892/ol.2018.8858
- Magadam, A., and Engel, F. B. (2018). Pparβ/Δ: Linking metabolism to regeneration. *Int. J. Mol. Sci.* 19 (7), 2013. doi:10.3390/ijms19072013
- Magnusson, K., Gremel, G., Rydén, L., Pontén, V., Uhlén, M., Dimberg, A., et al. (2016). Anln is a prognostic biomarker independent of ki-67 and essential for cell cycle progression in primary breast cancer. *BMC cancer* 16 (1), 904. doi:10.1186/s12885-016-2923-8
- Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biol.* 15 (6), 122. doi:10.1186/gb4184
- Mount, D. W., and Pandey, R. (2005). Using bioinformatics and genome analysis for new therapeutic interventions. *Mol. cancer Ther.* 4 (10), 1636–1643. doi:10.1158/1535-7163.mct-05-0150
- Olakowski, M., Tyszkiewicz, T., Jarzab, M., Król, R., Oczko-Wojciechowska, M., Kowalska, M., et al. (2009). Nbl1 and anillin (anln) genes over-expression in pancreatic carcinoma. *Folia Histochem. Cytobiol.* 47 (2), 249–255. doi:10.2478/v10042-009-0031-1
- Pai, V. C., Hsu, C. C., Chan, T. S., Liao, W. Y., Chuu, C. P., Chen, W. Y., et al. (2019). Aspm promotes prostate cancer stemness and progression by augmenting wnt-dvl-3-B-catenin signaling. *Oncogene* 38 (8), 1340–1353. doi:10.1038/s41388-018-0497-4
- Pujana, M. A., Han, J. D., Starita, L. M., Stevens, K. N., Tewari, M., Ahn, J. S., et al. (2007). Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* 39 (11), 1338–1349. doi:10.1038/ng.2007.2
- Qian, X., Song, X., He, Y., Yang, Z., Sun, T., Wang, J., et al. (2015). Ccnb2 overexpression is a poor prognostic biomarker in Chinese nscl patients. *Biomed. Pharmacother. = Biomedicine Pharmacother.* 74, 222–227. doi:10.1016/j.biopha.2015.08.004
- Rein, D. T., Roehrig, K., Schöndorf, T., Lazar, A., Fleisch, M., Niederacher, D., et al. (2003). Expression of the hyaluronan receptor rhamm in endometrial carcinomas suggests a role in tumour progression and metastasis. *J. cancer Res. Clin. Oncol.* 129 (3), 161–164. doi:10.1007/s00432-003-0415-0
- Rekhtman, N., Paik, P. K., Arcila, M. E., Tafe, L. J., Oxnard, G. R., Moreira, A. L., et al. (2012). Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: Lack of egfr/kras and presence of pik3ca/akt1 mutations. *Clin. cancer Res.* 18 (4), 1167–1176. doi:10.1158/1078-0432.ccr-11-2109
- Rouam, S., Moreau, T., and Broët, P. (2010). Identifying common prognostic factors in genomic cancer studies: A novel index for censored outcomes. *BMC Bioinforma.* 11, 150. doi:10.1186/1471-2105-11-150
- Saleh, A. A., Gohar, S. F., Hemida, A. S., Elgharabawy, M., and Soliman, S. E. (2020). Evaluation of aspm and tef gene expressions as potential biomarkers for bladder cancer. *Biochem. Genet.* 58 (3), 490–507. doi:10.1007/s10528-020-09962-1
- Sinha, D., Duijff, P. H. G., and Khanna, K. K. (2019). Mitotic slippage: An old tale with a new twist. *Cell cycle Georget. Tex* 18 (1), 7–15. doi:10.1080/15384101.2018.1559557

- Song, J. D., Lee, S. K., Kim, K. M., Park, S. E., Park, S. J., Kim, K. H., et al. (2009). Molecular mechanism of diallyl disulfide in cell cycle arrest and apoptosis in hct-116 colon cancer cells. *J. Biochem. Mol. Toxicol.* 23 (1), 71–79. doi:10.1002/jbt.20266
- Stella, G. M., Luisetti, M., Pozzi, E., and Comoglio, P. M. (2013). Oncogenes in non-small-cell lung cancer: Emerging connections and novel therapeutic dynamics. *Lancet Respir. Med.* 1 (3), 251–261. doi:10.1016/s2213-2600(13)70009-2
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). String V11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids Res.* 47 (1), D607–D13. doi:10.1093/nar/gky1131
- Tandon, D., and Banerjee, M. (2020). Centrosomal protein 55: A new paradigm in tumorigenesis. *Eur. J. Cell Biol.* 99 (5), 151086. doi:10.1016/j.ejcb.2020.151086
- Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2019). Five-year survival and correlates among patients with advanced melanoma, renal cell carcinoma, or non-small cell lung cancer treated with nivolumab. *JAMA Oncol.* 5 (10), 1411–1420. doi:10.1001/jamaoncol.2019.2187
- Varis, A., Salmela, A. L., and Kallio, M. J. (2006). Cenp-F (mitosin) is more than a mitotic marker. *Chromosoma* 115 (4), 288–295. doi:10.1007/s00412-005-0046-0
- Wang, A., Dai, H., Gong, Y., Zhang, C., Shu, J., Luo, Y., et al. (2019). Anln-induced Ezh2 upregulation promotes pancreatic cancer progression by mediating mir-218-5p/lasp1 signaling Axis. *J. Exp. Clin. cancer Res. CR* 38 (1), 347. doi:10.1186/s13046-019-1340-7
- Wong, D. W., Leung, E. L., So, K. K., Tam, I. Y., Sihoe, A. D., Cheng, L. C., et al. (2009). The eml4-alk fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type egfr and kras. *Cancer* 115 (8), 1723–1733. doi:10.1002/cncr.24181
- Wu, G., Zhou, L., Khidr, L., Guo, X. E., Kim, W., Lee, Y. M., et al. (2008). A novel role of the chromokinesin Kif4a in DNA damage response. *Cell cycleGeorget. Tex* 7 (13), 2013–2020. doi:10.4161/cc.7.13.6130
- Xu, J., Zheng, H., Yuan, S., Zhou, B., Zhao, W., Pan, Y., et al. (2019). Overexpression of anln in lung adenocarcinoma is associated with metastasis. *Thorac. cancer* 10 (8), 1702–1709. doi:10.1111/1759-7714.13135
- Yu, H. A., Tian, S. K., Drilon, A. E., Borsu, L., Riely, G. J., Arcila, M. E., et al. (2015). Acquired resistance of egfr-mutant lung cancer to a t790m-specific egfr inhibitor: Emergence of a third mutation (C797s) in the egfr tyrosine kinase domain. *JAMA Oncol.* 1 (7), 982–984. doi:10.1001/jamaoncol.2015.1066
- Yu, Y., Jiang, X., Schoch, B. S., Carroll, R. S., Black, P. M., and Johnson, M. D. (2007). Aberrant splicing of cyclin-dependent kinase-associated protein phosphatase kap increases proliferation and migration in glioblastoma. *Cancer Res.* 67 (1), 130–138. doi:10.1158/0008-5472.can-06-2478
- Zhang, H., Zhang, X., Li, X., Meng, W. B., Bai, Z. T., Rui, S. Z., et al. (2018). Effect of Ccnb1 silencing on cell cycle, senescence, and apoptosis through the P53 signaling pathway in pancreatic cancer. *J. Cell. physiology* 234 (1), 619–631. doi:10.1002/jcp.26816
- Zhang, L., and Maddox, A. S. (2010). Anillin. *Curr. Biol. CB* 20 (4), R135–R136. doi:10.1016/j.cub.2009.12.017
- Zhou, J. W., Wang, H., Sun, W., Han, N. N., and Chen, L. (2020). Aspm is a predictor of overall survival and has therapeutic potential in endometrial cancer. *Am. J. Transl. Res.* 12 (5), 1942–1953.
- Zou, Y., Watters, A., Cheng, N., Perry, C. E., Xu, K., Alicea, G. M., et al. (2019). Polyunsaturated fatty acids from astrocytes activate ppar signaling in cancer cells to promote brain metastasis. *Cancer Discov.* 9 (12), 1720–1735. doi:10.1158/2159-8290.cd-19-0270