Systematic Analysis Identifies NUF2 as an Immunological and Prognostic Biomarker for Non-small cell lung cancer

xia li
the sixth affiliated hospital of nantong university
zhongquan yi
he sixth affiliate hospital of nantong university□yancheng third people hospital
lianlian zhang
the fourth affiliated hospital of nantong university
jin zhou
the sixth affilated hospital of nantong university:yancheng third people hospital
wenchun song
the sixth affilated hospital of nantong university:yancheng tihrd people hospital
panwen zhao
Nantong Fifth People's Hospital: Nantong Tumor Hospital
jixiang wu
the sixth affilated hospital of nantong university:yancheng third people hospital
jianxiang song (jxsongycsy@163.com)
the sixth affilated hospital of nantong university:yancheng third people hospital
qinggan Ni
The fourth Affiliated Hospital of Nantong University

Research

Keywords: NUF2, Immunological, Non-small cell lung cancer, Prognostic, GEO

Posted Date: August 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-789591/v1

Systematic Analysis Identifies NUF2 as an Immunological and Prognostic Biomarker for Non-small cell lung cancer

Correspondence to: Jianxiang Song, jxsongycsy@163.com, Qinggan Ni,ycyy15171@163.com

Jianxiang Song, Department of Thoracic Surgery, The Sixth Affiliated Hospital of Nantong University：Yancheng Third People's Hospital, West Xindu Road 2#, Yancheng, Jiangsu Province, 224001, PR China. E-mail: jxsongycsy@163.com, Qinggan Ni，Department of Burns and Plastic Surgery, The fourth Affiliated Hospital of Nantong University：Yancheng First People's Hospital，West Yulong Road 166#, Yancheng, Jiangsu Province, 224001, PR China. E-mail: ycyy15171@163.com,

Xia Li[1,2*] ,Zhongquan Yi[2*], Lianlian zhang[3*], Jing Zhou[1],Wenchun Song[1],Panwen Zhao[2],Jixiang Wu[5],Jianxiang Song[5],Qinggan ni[4]

Abstract

Background: Non-small cell lung cancer (NSCLC) is a part of the most common cancers in the world. A lot of efforts have been made to clarify the etiology of non-small cell lung cancer, but the molecular mechanism of non-small cell lung cancer is still unclear.

Methods: In order to identify candidate genes in the occurrence and progression of non-small cell lung cancer, GSE19804 GSE118370 GSE19188 GSE27262 and GSE33532 microarray data sets were downloaded from the Gene Expression General (GEO) database. Identify differentially expressed genes (DEGS) and perform functional enrichment analysis. The protein-protein interaction network (PPI) was constructed, and the module analysis was performed using STRING and Cytoscape. A total of 562 DEGS were identified, consisting of 98 downregulated genes and 464 upregulated genes. Abundant functions and pathways of DEGS include p53 signaling pathway, Cell adhesion molecules,Leukocyte transendothelial migration,Vascular smooth muscle contraction,Complement and coagulation cascades

and Axon guidance. Tumor immunity was assessed to investigate the functions of hub genes.

Results: Totally 562 genes were found to be dysregulated. 12 genes were considered to be the hub genes. NUF2 was considered as the potential immunotherapeutic targets with future clinical significance.12 hub genes were detected, and biological process analysis showed that these genes are mainly enriched in p53 signaling pathway,Progesterone-mediated oocyte maturation,Cell cycle, Oocyte meiosis and Cellular senescence Survival analysis shows that NUF2 may be linked to the occurrence, invasion or recurrence of non-small cell lung cancer.

Conclusion: NUF2 genes discovered in this study help us include the molecular mechanisms of the occurrence and progression of non-small cell lung cancer, and provide candidate targets for the diagnosis and treatment of non-small cell lung cancer.

Keywords: NUF2, Immunological, Non-small cell lung cancer, Prognostic,GEO

Background

With the increase of life pressure, lifestyle changes, the decline of environmental quality and tobacco premature over-inhalation and a series of other reasons, the incidence of tumors has been high, of which, non-small cell lung cancer is very typical[1]. Non-small cell mortality rate is one of the highest in all malignant tumors. The five-year survival rate of non-small cell lung cancer is lower than other tumors[2].For ordinary families, once a family member develops non-small cell lung cancer, the blow to the family is huge. How to reduce the incidence of non-small cell lung cancer? First of all, we must have a healthy lifestyle, and secondly, high-end and precise treatment methods are essential. The development and progression of non-small cell lung cancer are shown to be linked to various factors, including genetic aberrations and immune infiltration[4]. Despite extensive studies on the mechanisms of carcinogenesis and progression, the etiology of non-small cell lung cancer is still unclear[3]. With the help of bioinformatics analysis tools and major database data, we can better find a target to combat tumors, and achieve early detection and prompt intervention in the early stages of tumors to avoid further development of tumors[5].

The histology of non-small cell lung cancer mainly includes lung adenocarcinoma, lung squamous cell carcinoma, large cell carcinoma and other types. Its development is multi-step, with abnormal genes as the main feature, aberrant gene expression that leads to phenotypic cell transformation[6-8]. Ribonuclease acid sequencing (RNA-Seq) has been used to detect genetic changes within the genome [9]. Comprehensive and systematic studies of the interaction between differential expression pathways and protein-coding genes can more accurately identify the carcinogenic effects of changes that occur during the progress and development of non-small cell lung cancer. Therefore, using these methods of bioinformatics to analyze RNA-Seq data can help understand the molecular pathogenesis and identify relevant tumor biomarkers[10]. RNA-Seq has been used to identify many key genes involved in disease progression that may help understand the gene expression changes that occur during the progress

of non-small cell lung cancer. So far, the principal driving force for carcinogenesis is still unclear, limiting the progress of targeted therapy for non-small cell lung cancer[11-14]. Therefore, understanding the pathogenesis of this disease is still a major challenge, and many key genes have not yet been identified.

In the past few decades, micro array technology and biotin morphology analysis have been widely used to screen genes changes at the genomic level that help us identify the functional carcinogenesis of differentially expressed genes (DEG) and the pathways involved in the development of non-small lung cancer[15]. However, the false positive rate in independent microarray analysis makes it difficult to obtain reliable results. Therefore, in this study, we first chose five gene sets (GSE19804[16] GSE118370[17] GSE19188[18] GSE27262[19] GSE33532[20]). Secondly, we used the R package from the Bioconductor project [21] and Venn's "LIMMA" graphic software to obtain differentially expressed sets of shared genes (DEG) in the above five data. Third, the database for annotation, visualization and comprehensive Discovery (DAVID) was analyzed. These DEGs include molecular functions (MF), cell components (CC), biological processes (BP) and Kyoto Protocol Encyclopedia of Genes and Genomes (KEGG) way. Fourth, we create a protein-protein interaction (PPI) network, and then applied cellular MCODE (Molecular Complexity Detection) to perform additional DEG analysis to determine some important modules. Fifth, MCODE is utilized to screen 12 Central genes. In addition, these dominant genes were imported online into the Kaplan Meier plotter database to obtain important prognostic information ($p < 0.05$). At the same time, We further verified that DEG, the hub between non-small cell lung cancer tissue and normal lung tissue, was tested by gene expression profiling (GEPIA; $P < 0.05$). All in all, the goal of this research is to improve the understanding of the carcinogenic effects of non-small cell lung cancer through the analysis of information about the process of genetic changes that occur during the disease and reveal the expression of central genes that can be used as biomarkers for clinical diagnosis, treatment and disease progression monitor.

Results

Identification of DEGs in NSCLC.

After standardization of the microarray results, DEGs were identified. The overlap among the 5 datasets contained 562 genes as shown in the Venn diagram (Fig. 1A), consisting of 98 downregulated genes and 464 upregulated genes between non-small cell lung cancer tissues and non-cancerous tissues.

KEGG and GO enrichment analyses of DEGs.

To analyze the biological classification of DEGs, functional and pathway enrichment analyses were performed using DAVID. GO analysis results showed that changes in biological processes (BP) of DEGs were significantly enriched in extracellular matrix organization、extracellular structure organization、nuclear division、mitotic nuclear division、organelle fission、cell-substrate adhesion、cell junction assembly、cell junction organization、vascular process in circulatory system、ameboidal-type cell migration (Table I). Changes in molecular function (MF) were mainly enriched in actin binding、amyloid-beta binding、extracellular matrix structural constituent、peptide binding、amide binding、histone kinase activity、metalloendopeptidase activity、extracellular matrix structural constituent conferring tensile strength、metallopeptidase activity (Table I). Changes in cell component (CC) of DEGs were mainly enriched in thecell-cell junction、actin filament bundle、stress fiber、contractile actin filament bundle、contractile fiber part、midbody、condensed chromosome, centromeric region、chromosomal region、condensed nuclear chromosome (Table I). KEGG pathway analysis revealed that the downregulated DEGs were mainly enriched in p53 signaling pathway, while the upregulated DEGs were mainly enriched in Cell adhesion molecules、Leukocyte transendothelial migration、Vascular smooth muscle contraction、Complement and coagulation cascades and Axon guidance.

PPI network construction and module analysis.

The PPI network of DEGs was constructed (Fig. 1C) and the most significant module was obtained using Cytoscape (Fig. 1B). The functional analyses of genes involved in this module were analyzed using DAVID. Results showed that genes in this module were mainly enriched in nuclear division、organelle fission、mitotic nuclear division、cell cycle checkpoint 、 histone phosphorylation 、 condensed chromosome 、 chromosomal region 、 condensed chromosome, centromeric region 、 midbody 、 chromosome, centromeric region 、 histone kinase activity 、 protein serine/threonine kinase activity 、 protein C-terminus binding 、 ferric iron binding 、 oxidoreductase activity, acting on CH or CH2 groups、p53 signaling pathway、Progesterone-mediated oocyte maturation 、Cell cycle、Oocyte meiosis and Cellular senescence. (Table II).

Hub gene selection and analysis.

A total of 12 genes were identified as hub genes with degrees ≥10. The names, abbreviations and functions for these hub genes are shown in （Table III）.The 12 hub genes were used to draw the difference in the distribution of lung adenocarcinoma and lung squamous cell carcinoma tissues and adjacent tissues in the TCGA (https://portal.gdc.cancer.gov/) database using ggplot2 in R language.Tthe expression of hub gene in cancer tissues and adjacent tissues of unpaired samples in LUAD（Fig. 2A）, The expression of hub gene in cancer tissues and adjacent tissues of unpaired samples in LUSD（Fig. 2B）. The expression of hub gene in LUAD paired sample cancer tissues and adjacent tissues（Fig. 2C）,The expression of hub gene in LUSD paired sample cancer tissues and adjacent tissues（Fig. 2D）. The P value adopts scientific notation and all have statistical significance.$p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Download the hub gene from the DAIVED website (https://david.ncifcrf.gov/), and then use the ggplot2 package and clusterProfiler package in the R language to visualize the enrichment analysis of GO and KEGG. （Fig. 3A）is the enrichment analysis of up-regulated genes, and（Fig. 3B） is the enrichment analysis of down-regulated genes. （Fig. 3C）is the visualization diagram of the enrichment analysis of the hub gene. The 12 hub genes were analyzed with the

pROC package in R language and visualized with the ggplot2 package.（Fig. 4A、4B）are receiver operating characteristic curves (ROC) of LUAD, and（Fig. 4C、4D） are receiver operating characteristic curves (ROC) of LUSC. It can be seen that whether it is in lung adenocarcinoma or lung squamous cell carcinoma, the predictive ability of variable NUF2 has higher accuracy in predicting the outcome of Normal and Tumor. Subsequently, the overall survival analysis of the hub genes was performed using Kaplan-Meier curve. NSCLC patients with CDK1、 CCNB1、 TOP2A、 PRM2、 CHEK1、AURKA、ZWINT、NUF2、MKI67、BIRC5、CEP55、ANLN showed worse overall survival (Fig. 5A-H). We noticed that NSCLC patients the association of changes in the NUF2 genome shows a decrease in overall survival, while NSCLC patients NUF2 genome changes show the highest hazard ratio. The observations are statistically significant（HR＝2.01 CI1.7-2.39 P＝2.4e-16）(Fig. 5H). The expression of NUF2 mRNA in a variety of cancer types shows that the expression of NUF2 mRNA in cancer tissues is higher than that in adjacent tissues: (Fig 6A) Comparison shows that NUF2 mRNA is overexpressed (left column, red) and underexpressed (right column, blue) The data set is in cancer and normal tissues. The diagram comes from the Oncomine database (available from https://www.oncomine.org/resource/login.html). The threshold uses the following parameters: p value is 1E-4, fold change is 2, and gene ranking is 10%. (Fig 6B) Expression of NUF2 in 33 human cancers (Gene Expression Profiling Interactive Analysis 2) (URL https://gepia2.cancer-pku.cn) Obtained from Cancer Genome Atlas via GEPIA2: Dot plot showing all gene expression profiles The tumor samples and matched normal tissues were collected. Each point represents the expression of the sample. (Fig 6C) The expression pattern of NUF2 mRNA in tumors and tissues. Corresponding normal tissue: From the GENT (gene expression in normal and tumor tissues) database (http://medicalgenomics.kribb.re.kr/GENT/). The boxes represent the median and the 25th and 75th percentiles. Dots represent outliers. The red box represents tumor tissue, and the blue box represents normal tissue.

Use proteinatlas network database (https://www.proteinatlas.org/) to analyze: Fig 7A is the HE staining result of NUF2 in lung adenocarcinoma, and Fig 7B is the HE

staining result of NUF2 in lung squamous cell carcinoma. Fig 7C is the HE staining result of NUF2 in normal bronchus. Fig 7D is the use of SAGE to analyze NUF2 in human cancers.The TIMER website (https://cistrome.shinyapps.io/timer/) was used to show the correlation between tumor immune cell infiltration and NUF2 somatic copy number in LUAD and LUSD, respectively. According to the copy number of NUF2, the samples were divided into five categories (Deep Deletion, Arm-level Deletion, Diploid/Normal, Arm-level Gain, High Amplication), and the distribution of infiltrated immune cells in the five samples was compared. The display result is: In LUAD, except for the significant difference in the copy number of Arm-level Deletion in B cell, the other five types of CD8+T cell, CD4+T cell, Macrophage, Neutrophils, and Dendritic cell are Diploid/Normal, Arm- There is a significant difference in level Gain. However, the LUSC is not exactly the same. The copy numbers in B cell, Neutrophils, and Dendritic cells have obvious differences. The copy numbers of Arm-level Deletion, Arm-level Gain, and High Amplication in CD4+T cells have obvious differences. Arm-level in Macrophage Gain has obvious difference, and any sample difference in CD8+ cell is statistically significant. The code of P value is significant: $0 \leq *** < 0.001 \leq ** < 0.01 \leq * < 0.05$ (Fig 8A). Use the TIMER website (https://cistrome.shinyapps.io/timer/) to view the correlation of immune cells and tumor purity and NUF2 expression in LUAD and LUSD in the TCGA database(Fig 8B). Table 4 shows the Correlation analysis between NUF2 and immune cell-related genes and markers in TIMER $0 \leq *** < 0.001 \leq ** < 0.01 \leq * < 0.05$. Use the UALCAN website (http://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl) respectively shows the expression differences of NUF2 lung adenocarcinoma tumors in different grades, normal population and lung adenocarcinoma The expression of NUF2 in different smokers and the difference in the expression of NUF2 in TP53 mutant and non-mutated lung adenocarcinoma (Fig 9A-C). The differences in the expression of NUF2 lung squamous cell carcinoma tumors of different grades, the expression of NUF2 in normal people and different smokers of lung squamous cell carcinoma, and the difference in expression of NUF2 in TP53 mutated and non-mutated lung squamous cell carcinoma. *** indicates that the difference is

statistically significant(Fig 9D-F).

Discussion

Lung cancer is the most common cause of cancer death in the world.An estimated 1.6 million people die each year[35]， Approximately 85% of lung cancer patients have non-small cell lung cancer pathological types, among which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes of non-small cell lung cancer[36]. With the continuous development of molecular biology and information technology, in the past two decades, we have made significant progress and progress in the treatment of non-small cell lung cancer (NSCLC)[37].The most common etiology of lung cancer is smoking, and it is also related to environmental exposure, including second-hand smoking, pollution and occupational carcinogens, as well as genetic susceptibility, etc[38,39]. However, the specific mechanism of the occurrence and development of non-small cell lung cancer is not extremely clear. Cell cycle regulators play deeply influential roles in NSCLC[40]. MiR-34b-3p represses cell proliferation, cell cycle progression and cell apoptosis in non-small-cell lung cancer (NSCLC) by targeting CDK4[41], Chidamide-induced ROS accumulation and miR-129-3p-dependent cell cycle arrest in non-small lung cancer cells42. Most NSCLC cases that are not detected early are not candidates for treatment, which may be one of the reasons for the poor prognosis of patients. Therefore, there is an urgent need for potential markers for efficient diagnosis and treatment. Microarray technology allows us to explore genetic changes in NSCLC and has proven to be a useful method for identifying new biomarkers in other diseases43.

In this study, five mRNA microarray data sets were analyzed to obtain DEG between non-small cell lung cancer tissues and non-cancerous tissues. A total of 562 DEGs were identified in five data sets, including 98 down-regulated genes and 464 up-regulated genes. GO and KEGG enrichment analysis was performed to explore the interaction between DEG.KEGG pathway analysis revealed that the downregulated DEGs were mostly enriched in p53 signaling pathway, while the upregulated DEGs

were mostly enriched in Cell adhesion molecules、Leukocyte transendothelial migration、Vascular smooth muscle contraction、Complement and coagulation cascades and Axon guidance. these are closely related to immune infiltration，Fathia Mami-Chouaib et al. studied Resident that memory T cells, critical components in tumor immunology44.The tumor microenvironment will affect the disease progression of many human malignant tumors. Infiltration of immune-related cells into tumors will increase the recruitment of immune activation and anti-disease immune effector cells and pathways45. In recent years, studies have shown that p53 protein can mediate the nucleolar stress responses, leading to cell cycle arrest, apoptosis, senescence or differentiation, thereby affecting the occurrence and development of tumors46. At the same time, the same time, studies have clarified that mutations of the p53 tumor suppressor gene often occur in lung cancer. Mutant p53 (mtp53) can inhibit wild-type p53 protein and destroy its tumor suppressor function. In addition, mutant p53 usually functions as an oncogene. The post-translational modification of p53 protein is important for its transcription and tumor suppressor function47. In short, short, all these theories are consistent with our results. GO analysis results showed that DEGs were significantly enriched in extracellular matrix organization、extracellular structure organization、nuclear division、mitotic nuclear division、organelle fission、cell-substrate adhesion、cell junction assembly、cell junction organization、vascular process in circulatory system、ameboidal-type cell migration,KEGG pathway analysis revealed that DEGs were mainly enriched in p53 signaling pathway, Cell adhesion molecules、Leukocyte transendothelial migration、Vascular smooth muscle contraction、Complement and coagulation cascades and Axon guidance.

We selected 12 DEGs as hub genes with degrees ≥10. Among these hub genes, We noticed that NSCLC patients the association of changes in the NUF2 genome shows a decrease in overall survival, while NSCLC patients NUF2 genome changes show the highest hazard ratio.NUF 2, which acts as a component of the essential kinetochore- associated NDC80 complex, which is required for chromosome segregation and spindle checkpoint activity. Required for kinetochore integrity and

the organization of stable microtubule binding sites in the outer plate of the kinetochore. The NDC80 complex synergistically enhances the affinity of the SKA1 complex for microtubules and may allow the NDC80 complex to track depolymerizing microtubules48. NUF2 is reported as one of the cancer testis antigens, which is ectopically secreted by cancer, and the expression level of NUF2 is elevated in prostate cancer tissues49. Xingwei Xie et al. found that Nuf2 participates in the regulation of cell apoptosis and proliferation by regulating the binding of centromere and spindle microtubules to achieve the correct separation of chromosomes. And Nuf2 Is a Prognostic-Related Biomarker and Correlated With Immune Infiltrates in Hepatocellular Carcinoma50. In addition, addition, NUF2 is overexpressed in breast cancer, human osteosarcoma, pancreatic cancer, colorectal cancer, and may be regarded as a valuable biomarker for diagnosis, treatment and prognosis of tumors51-54. These indicate that NUF2 can be used as an effective indicator for predicting the prognosis of NSCLC.

Tumor microenvironment (TME) is a key regulator of tumorigenesis, progression and resistance, and is closely related to tumor drug resistance mechanisms55. In the tumor microenvironment (TME), tumor, cells continue to evolve to reduce the production of new antigens and the burden of mutations to evade anti-tumor responses. This will reduce the tumor's responsiveness to the adaptive immune response and produce tumor internal factors, such as changes in the expression of immunomodulatory molecules on tumor cells. External tumor factors, such as immunosuppressive cells, soluble inhibitory molecules, or inhibitory receptors expressed by immune cells, can change the composition and activity of tumor infiltrating lymphocytes (TIL) (by increasing the T regulatory cell: T effector cell ratio and suppressing T Effector cell function) and promote tumor growth and metastasis56. Through the analysis of this study, NUF2 is related to immune infiltration classification in NSCLC, such as CD8+T cell, T cell (general), B cell, Monocyte, TAM, M1 Macrophage, M2 Macrophage, Neutrophils, Natural killer cell, Dendritic cell, Th1, Th2, Tfh, Th17, Treg; T cell exhaustions are all related. The correlation between NUF2 and immunosuppressive gene expression indicates that

NUF2 It plays a key role in regulating tumor immunology.

In summary, this study aims to identify DEGs that may be involved in the carcinogenesis or progression of NSCLC. A total of 562 DEG and 12 central genes were identified, which can be used as diagnostic biomarkers for NSCLC. At the same time, it further proves that NUF2 can be used as an effective immunotherapy target. In the next step, our research group will use molecular biology experimental methods to further verify the biological functions of NUF2 in NSCLC from in vivo and in vitro experiments.

Methods

Microarray data.

GEO (http://www.ncbi.nlm.nih.gov/geo)[22] is a public functional genomics data repository of high throughout gene expression data, chips and microarrays. Five gene expression datasets [GSE19804[16] GSE118370[17] GSE19188[18] GSE27262[19] GSE33532[20]] were downloaded from GEO (Affymetrix GPL570 platform Affymetrix Human Genome U133 Plus 2.0 Array respectively,The probes were converted into the corresponding gene symbol according to the annotation information in the platform. The GSE19804 dataset contained 60 Non-small cell lung cancer tissue samples and 60 non-cancerous samples. The GSE118370 dataset contained 6 Non-small cell lung cancer tissue samples and 6 non-cancerous samples. The GSE19188 dataset contained 91 Non-small cell lung cancer tissue samples and 65 non-cancerous samples. The GSE27262 dataset contained 25 Non-small cell lung cancer tissue samples and 25 non-cancerous samples.The GSE33532 dataset contained 80 Non-small cell lung cancer tissue samples and 20 non-cancerous samples.

Identification DEGS.

DEGS between non-small cell lung cancer and non-cancer samples were screened by GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r). GEO2R is a cross - activity Web tool, Allows users to compare two or more data sets in GEO series, For identifying

DEGS. under experimental conditions P value (adj.) adjusted P) and Benjamini and Hochberg false discovery rates, provides a balance between the limitations of finding statistically significant genes and false positives. probe sets without corresponding gene symbols or genes with multiple probe sets were removed or averaged, respectively. Log FC(folding changes)>1 and adj. P value <0.01 was considered statistically significant.

DEG KEGG and GO enrichment analysis.

Database Discovery of annotation, visualization and integration (DAVID;http://david.ncifcrf.gov)(version6.7)[23] is an online version Biological Information Database for Integration of Biological Information Data and analysis tools and a comprehensive set of tools Functional annotation of genes and proteins used Users can extract biological information. KEGG is a database Resources for understanding advanced and biological functions Systems generated from large-scale molecular datasets high throughput experimental techniques[24.] GO is a professional course.

PPI network construction and module analysis.

The PPI network was predicted using Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 10.0)[25] online database. Analyzing the functional interactions between proteins may provide insights into the mechanisms of generation or development of diseases. In the present study, PPI network of DEGs was constructed using STRING database, and an interaction with a combined score >0.4 was considered statistically significant. Cytoscape (version 3.4.0) is an open source bioinformatics software platform for visualizing molecular interaction networks[26]. The plug-in Molecular Complex Detection (MCODE) (version 1.4.2) of Cytoscape is an APP for clustering a given network based on topology to find densely connected regions[27]. The PPI networks were drawn using Cytoscape and the most significant module in the PPI networks was identified using MCODE. The criteria for selection were as follows: MCODE scores >5, degree cut-off=2, node score

cut-off=0.2, Max depth=100 and k-score=2. Subsequently, the KEGG and GO analyses for genes in this module were performed using DAVID.

Hub genes selection and analysis.

The hub genes were selected Hub genes selection and analysis. The hub genes were selected with degrees ≥10. A network of the genes and their co-expression genes was analyzed using cBioPortal (http://www.cbioportal. org) [28,29] online platform. The biological process analysis of hub genes was performed and visualized using Biological Networks Gene Oncology tool (BiNGO) (version 3.0.3) plugin of Cytoscape[30] . Hierarchical clustering of hub genes was constructed using UCSC Cancer Genomics Browser (http:// genome-cancer.ucsc.edu) [31]. The overall survival and disease-free survival analyses of hub genes were performed using Kaplan-Meier curve in cBioPortal. The expression profiles of NUF2 was analyzed and displayed using online database Serial Analysis of Gene Expression (SAGE; http://www. ncbi.nlm.nih.gov/SAGE). The relationship between expression patterns and tumor grades, infection status, satellites and vascular invasion were analyzed using online database Oncomine (http://www.oncomine.com)[32-34] .

Abbreviations

Not applicable.

Acknowledgements

Not applicable.

Authors' contributions

Xia Li , Zhongquan Yi, Lianlian zhang researched and analyzed data. Wenchun Song, Panwen Zhao, Jixiang Wu contributed to the discussion. Jianxiang Song and Qinggan ni designed the study and Xia Li wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported in part by grant from the 2020 Clinical College (Yancheng Third People's Hospital) Special Scientific Research and Development Fund Project(20209113)

Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

We assure that the material is original and it has not been published elsewhere yet.

Competing interests

The authors declared that no competing of interests existing in this study.

Author details

1Department of General Medicine, The Sixth Affiliated Hospital of Nantong University：Yancheng Third People's Hospital , Jiangsu Province, China
2The Central Laboratory， The Sixth Affiliated Hospital of Nantong University：Yancheng Third People's Hospital，Jiangsu Province, China
3Department of Ultrasound Imaging, The fourth Affiliated Hospital of Nantong University：Yancheng First People's Hospital, Jiangsu Province, China
4Department of Burns and Plastic Surgery, The fourth Affiliated Hospital of Nantong University：Yancheng First People's Hospital，Jiangsu Province, China
5Department of Thoracic Surgery, The fourth Affiliated Hospital of Nantong UniversityYancheng Third People's Hospital,Jiangsu Province, China
*Equal contribution
Correspondence to: Jianxiang Song, Qinggan Ni,; email:jxsongycsy@163.com , ycyy15171@163.com,
Jianxiang Song, Department of Thoracic Surgery, The Sixth Affiliated Hospital of Nantong University： Yancheng Third People's Hospital, West Xindu Road 2#, Yancheng, Jiangsu Province, 224001, PR China. E-mail: jxsongycsy@163.com, Telephone: 86-15189200233, Fax: 86-051581600100
Qinggan Ni，Department of Burns and Plastic Surgery, The fourth Affiliated Hospital of Nantong University：Yancheng First People's Hospital，West Yulong Road 166#, Yancheng, Jiangsu Province, 224001, PR China. E-mail: ycyy15171@163.com, Telephone: 86-19851550550, Fax: 86-051588508845

References

1.Lais Osmani, Frederic Askin, Edward Gabrielson, Qing Kay Li,Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy.Semin Cancer Biol. 2018 Oct;52(Pt 1):103-109.

2.Elahe Khaksar,Mohsen Askarishahi, Seyedhossein Hekmatimoghaddam, Hassanali Vahedian-Ardakani,Cox Regression and Parametric Models: Comparison of How They Determine Factors Influencing Survival of Patients with Non-Small Cell Lung Carcinoma.Asian Pac J Cancer Prev. 2017 Dec 29;18(12):3389-3393.

3.Takayuki Kishikawa, Takashi Kasai, Masahiko Okada,Osimertinib, a third-generation EGFR tyrosine kinase inhibitor: A retrospective multicenter study of its real-world efficacy and safety in advanced/recurrent non-small cell lung carcinoma.Thorac Cancer. 2020 Apr;11(4):935-942.

4.G Boulle, Y Velut, A Mansuet-Lupo,Chemoradiotherapy efficacy is predicted by intra-tumour CD8+/FoxP3+ double positive T cell density in locally advanced N2 non-small-cell lung carcinoma.Eur J Cancer. 2020 Aug;135:221-229.

5.Lin Li, Qingsong Lei, Shujun Zhang,Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis.ONCOLOGY REPORTS 38: 2607-2618, 2017

6.G Boulle, Y Velut, A Mansuet-Lupo, L Gibault ,Chemoradiotherapy efficacy is predicted by intra-tumour CD8+/FoxP3+ double positive T cell density in locally advanced N2 non-small-cell lung carcinoma.Eur J Cancer. 2020 Aug;135:221-229

7.Yohan Bossé, Christopher I Amos,A Decade of GWAS Results in Lung Cancer.Cancer Epidemiol Biomarkers Prev. 2018 Apr;27(4):363-379.

8.Matthew B Schabath, Michele L Cote,Cancer Progress and Priorities: Lung Cancer.Cancer Epidemiol Biomarkers Prev. 2019 Oct;28(10):1563-1579.

9.Edgar R, Domrachev M and Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210, 2002.

10.Sudhanshu Shukla, Joseph R Evans, Rohit Malik, et al,Development of a RNA-Seq Based Prognostic Signature in Lung Adenocarcinoma.J Natl Cancer Inst. 2016 Oct 5;109(1):djw200.

11.Narjust Duma, Rafael Santana-Davila, Julian R Molina,Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment.Mayo Clin Proc. 2019 Aug;94(8):1623-1640.

12.Sushma Jonna, Deepa S Subramaniam,Molecular diagnostics and targeted therapies in non-small cell lung cancer (NSCLC): an update.Discov Med. 2019 Mar;27(148):167-170.

13.Lais Osmani, Frederic Askin, Edward Gabrielson, Qing Kay Li,Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy.Semin Cancer Biol. 2018 Oct;52(Pt 1):103-109.

14.Evan C Naylor, Jatin K Desani, Paul K Chung,Targeted Therapy and Immunotherapy for Lung Cancer.Surg Oncol Clin N Am. 2016 Jul;25(3):601-9.

15.Birgitte Sandfeld-Paulsen, Kristine Raaby Jakobsen, Rikke Bæk ,Exosomal Proteins as Diagnostic Biomarkers in Lung Cancer.J Thorac Oncol. 2016 Oct;11(10):1701-10.

16.Lu TP, Tsai MH, Lee JM, Hsu CP et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* 2010 Oct;19(10):2590-7.

17.Xu L, Lu C, Huang Y, Zhou J et al. SPINK1 promotes cell growth and metastasis of lung adenocarcinoma and acts as a novel prognostic biomarker. *BMB Rep* 2018 Dec;51(12):648-653.

18.Hou J, Aerts J, den Hamer B, van Ijcken W et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* 2010 Apr 22;5(4):e10312.

19.Wei TY, Juan CC, Hisa JY, Su LJ et al. Protein arginine methyltransferase 5 is a potential oncoprotein that upregulates G1 cyclins/cyclin-dependent kinases and the phosphoinositide 3-kinase/AKT signaling cascade. *Cancer Sci* 2012 Sep;103(9):1640-50.

20.Michael Meister, Anton Belousov, Xu EC, Schnabel P, Warth A, Hoffmann H, Dienemann H, Riedlinger J, Bodenmueller H, Zolg W, Herth FJF, Muley T. Intra-tumor Heterogeneity of Gene Expression Profiles in Early Stage Non-Small Cell Lung Cancer. Journal of Bioinformatics Research Studies 2014, 1(1): 1

21.Mok SC, Bonome T, Vathipadiekal V, Bell A, Johnson ME, Wong KK, Park DC, Hao K, Yip DK, Donninger H, *et al*: A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor:Microfibril-associated glycoprotein 2. Cancer Cell 16: 521-532, 2009.

22.Edgar R, Domrachev M and Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210, 2002.
23.Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki RA: The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8: R183, 2007.
24.Kanehisa M: The KEGG database. Novartis Found Symp 247: 91-252, 2002.
25.Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, *et al*: STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808-D815, 2013.
26.Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27: 431-432, 2011.
27.Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY and Arai AE: MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: A clinical validation study. J Cardiovasc Magn Reson 14: 83, 2012.
28.Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6: pl1, 2013.
29. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, *et al*: The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov 2: 401-404, 2012.
30.Maere S, Heymans K and Kuiper M: *BiNGO*: A Cytoscape plugin to assess overrepresentation of gene ontology categories 19. in biological networks. Bioinformatics 21: 3448-3449, 2005.
31.Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D: The human genome browser at UCSC. Genome Res 12: 996-1006, 2002.
32.Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, Lai KM, Ji J, Dudoit S, Ng IO, *et al*: Gene expression patterns in human liver cancers. Mol Biol Cell 13: 1929-1939, 2002.
33. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX, *et al*: A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 70: 10202-10212, 2010.
34. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, *et al*: Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology 45: 938-947, 2007.

35.Torre, L. A. et al. Global cancer statistics, 2012. CA Cancer J. Clin. 2015 Mar;65(2):87-108
36.Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* 2008 May;83(5):584-94.
37.Roy S. Herbst1, Daniel Morgensztern & Chris Boshoff, The biology and management of non-small cell lung cancer.Nature. 2018 Jan 24;553(7689):446-454.
38.Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* ,2013 May;143(5 Suppl):e1S-e29S.

39.Vineis, P. *et al.* Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *Br. Med. J.* 2005 Feb 5;330(7486):277.
40.Yawen Qiang , Fan Ma, Ziran Wang. *et al.*LukS-PV induces cell cycle arrest and apoptosis through p38/ERK MAPK signaling pathway in NSCLC cells,Biochem Biophys Res Commun. 2020 Jan 22;521(4):846-852.
41.Hongxiang Feng,*et al.*MiR-34b-3p represses cell proliferation, cell cycle progression and cell apoptosis in non-small-cell lung cancer (NSCLC) by targeting CDK4. J Cell Mol Med. 2019 Aug;23(8):5282-5291.
42.Yueh-Feng Wu,*et al.*Chidamide-induced ROS accumulation and miR-129-3p-dependent cell cycle arrest in non-small lung cancer cells Phytomedicine. 2019 Mar 15;56:94-102.
43.Lin Li, Qingsong Lei,*et al.*Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis,ONCOLOGY REPORTS 38: 2607-2618, 2017
44.Fathia Mami-Chouaib, Charlotte Blanc , Stéphanie Corgnac，*et al.*Resident memory T cells, critical components in tumor immunology，J Immunother Cancer. 2018 Sep 4;6(1):87.
45.Jo Marie Tran Janco 1, Purushottam Lamichhane 2, Lavakumar Karyampudi 3, Keith L Knutson，Tumor-infiltrating dendritic cells in cancer pathogenesis，J Immunol. 2015 Apr 1;194(7):2985-91.
46.Emiliano Matos-Perdomo, Félix Machín, Nucleolar and Ribosomal DNA Structure under Stress: Yeast Lessons for Aging and Cancer.Cell. 2019 Jul 26;8(8):779.
47.Xinyu Tang, Yang Li, Long Liu,*et al.*Sirtuin 3 induces apoptosis and necroptosis by regulating mutant p53 expression in small- cell lung cancer.Oncol Rep. 2020 Feb;43(2):591-600.
48.Divya Subramonian,Te-An Chen, Nicholas Paolini, Xiang-Dong David Zhang,Poly-SUMO-2/3 chain modification of Nuf2 facilitates CENP-E kinetochore localization and chromosome congression during mitosis.Cell Cycle. 2021 May;20(9):855-873.
49.Shinichiro Yamamoto, Ken-ichi Takayama, Daisuke Obinata, *et al.*Identification of new octamer transcription factor 1-target genes upregulated in castration-resistant prostate cancer. Cancer Sci. 2019 Nov; 110(11): 3476–3485.
50.Xingwei Xie, Shanshan Jiang, Xiang Li, *Nuf2* Is a Prognostic-Related Biomarker and Correlated With Immune Infiltrates in Hepatocellular Carcinoma,Front Oncol. 2021 Mar 9;11:621373.
51.Wenjie Xu, Yizhen Wang, Yanan Wang,et *al.*Screening of differentially expressed genes and identification of NUF2 as a prognostic marker in breast cancer.Int J Mol Med. 2019 Aug;44(2):390-404.
52.H-L Fu, L Shao,Silencing of NUF2 inhibits proliferation of human osteosarcoma Saos-2 cells.Eur Rev Med Pharmacol Sci. 2016;20(6):1071-9.
53.Peng Hu, Xi Chen, Jing Sun,et *al, S*iRNA-mediated knockdown against NUF2 suppresses pancreatic cancer proliferation in vitro and in vivo.Biosci Rep. 2015 Jan 14;35(1):e00170.
54.Hironobu Sugimasa, Kenzui Taniue, Akiko Kurimoto,et *al,*Heterogeneous nuclear ribonucleoprotein K upregulates the kinetochore complex component NUF2 and promotes the tumorigenicity of colon cancer cells.Biochem Biophys Res Commun. 2015 Mar 27;459(1):29-35.
55.Peijie Wu, Wei Gao, Miao Su,et *al,*Adaptive Mechanisms of Tumor Therapy Resistance Driven by Tumor Microenvironment.Front Cell Dev Biol. 2021 Mar 1;9:641469.
56.Reem Saleh, Eyad Elkord,Acquired resistance to cancer immunotherapy: Role of tumor-mediated immunosuppression.Semin Cancer Biol. 2020 Oct;65:13-27.

Figures

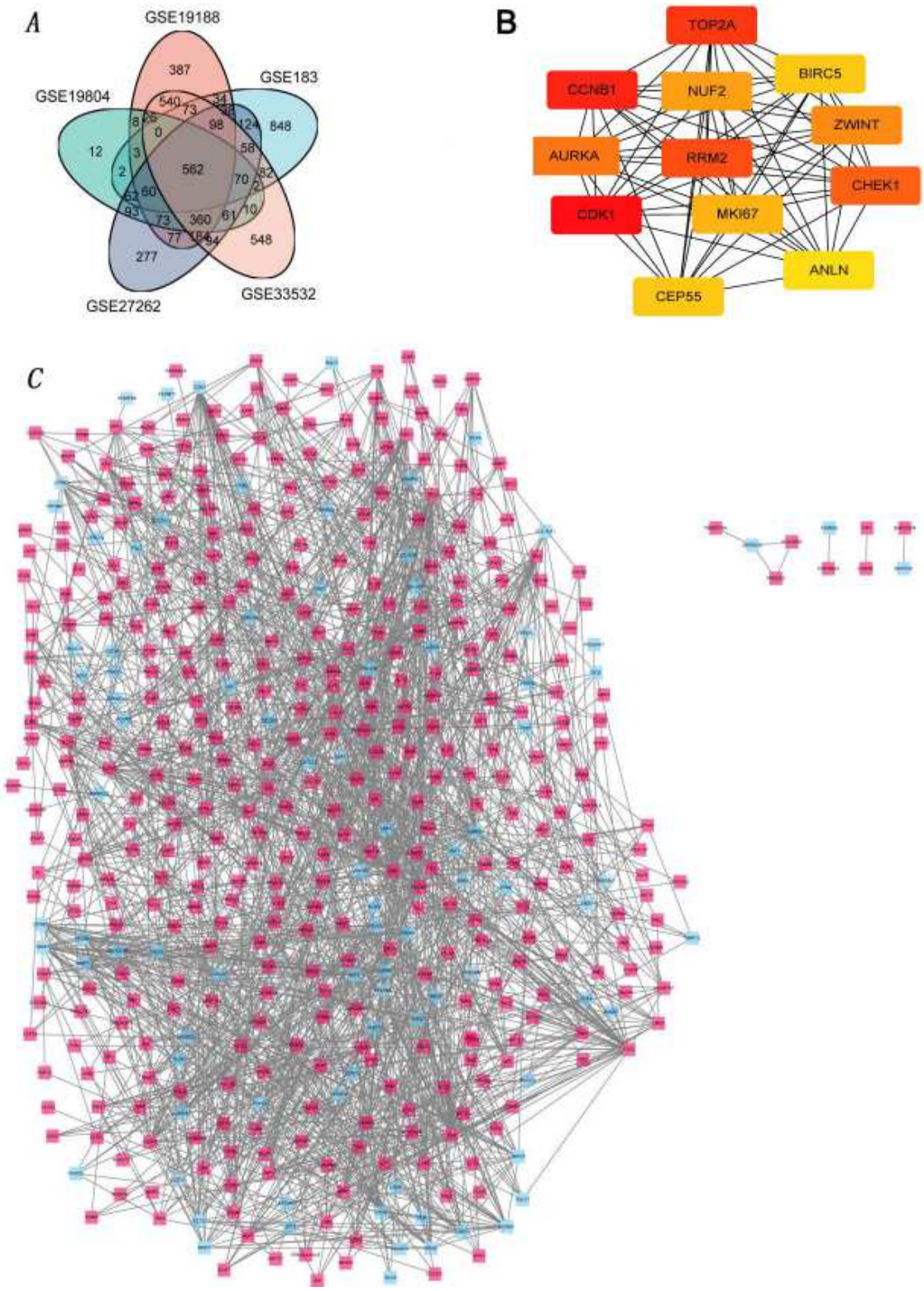

Figure 1

Venn diagram, PPI network and the most significant module of DEGs. (A) DEGs were selected with a fold change >1 and P-value <0.01 among the mRNA expression profiling set GSE19804 GSE118370 GSE19188 GSE27262 GSE33532. The 5 datasets showed an overlap of 562 genes. (B) The most

significant module was obtained from PPI network with 12 nodes and 66 edges. (C)The PPI network of DEGs was constructed using Cytoscape. Upregulated genes are marked in light red; downregulated genes are marked in light blue.

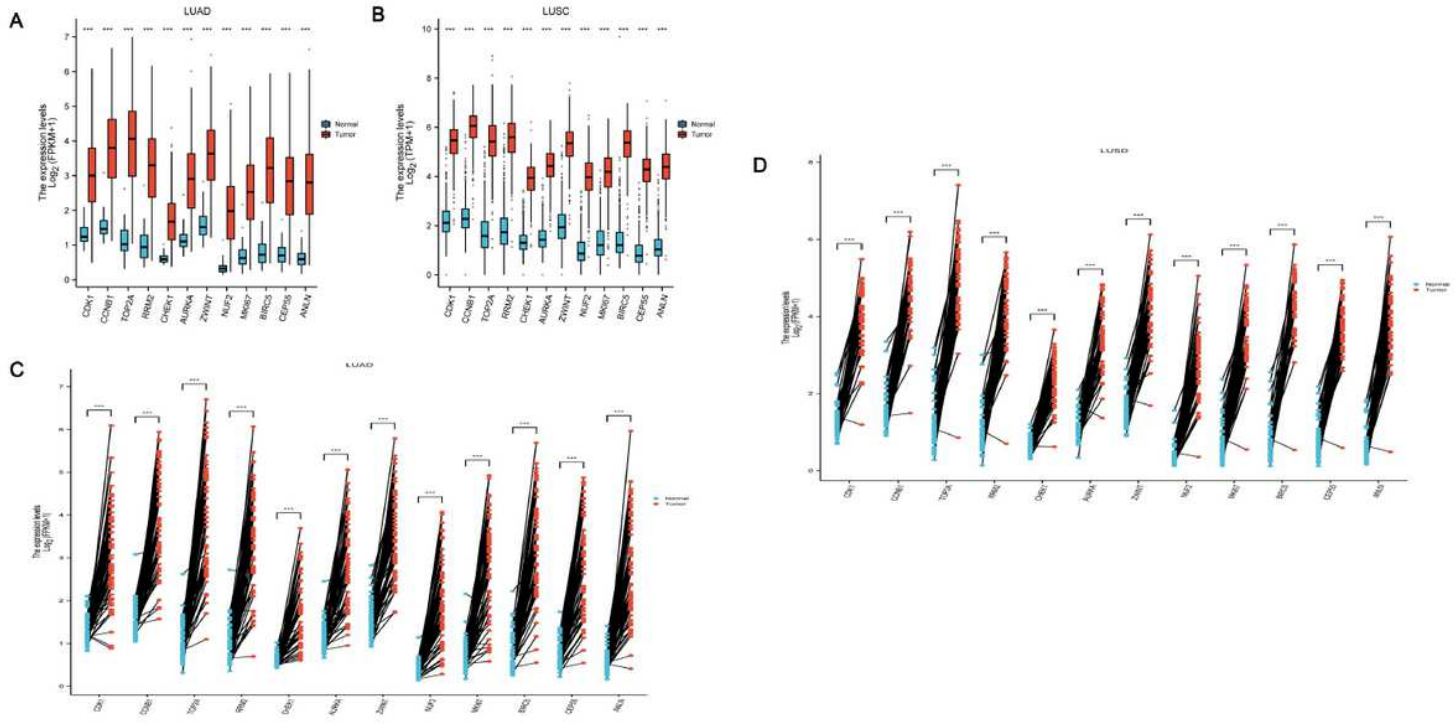

Figure 2

The 12 hub genes were used to draw the difference in the distribution of lung adenocarcinoma and lung squamous cell carcinoma tissues and adjacent tissues in the TCGA (https://portal.gdc.cancer.gov/) database using ggplot2 in R language. A is the expression of hub gene in cancer tissues and adjacent tissues of unpaired samples in LUAD, and B is the expression of hub gene in cancer tissues and adjacent tissues of unpaired samples in LUSD. C represents the expression of hub gene in LUAD paired sample cancer tissues and adjacent tissues, and D represents the expression of hub gene in LUSD paired sample cancer tissues and adjacent tissues. The P value adopts scientific notation and all have statistical significance. *, p< 0.05; **, p<0.01; ***, p<0.001.

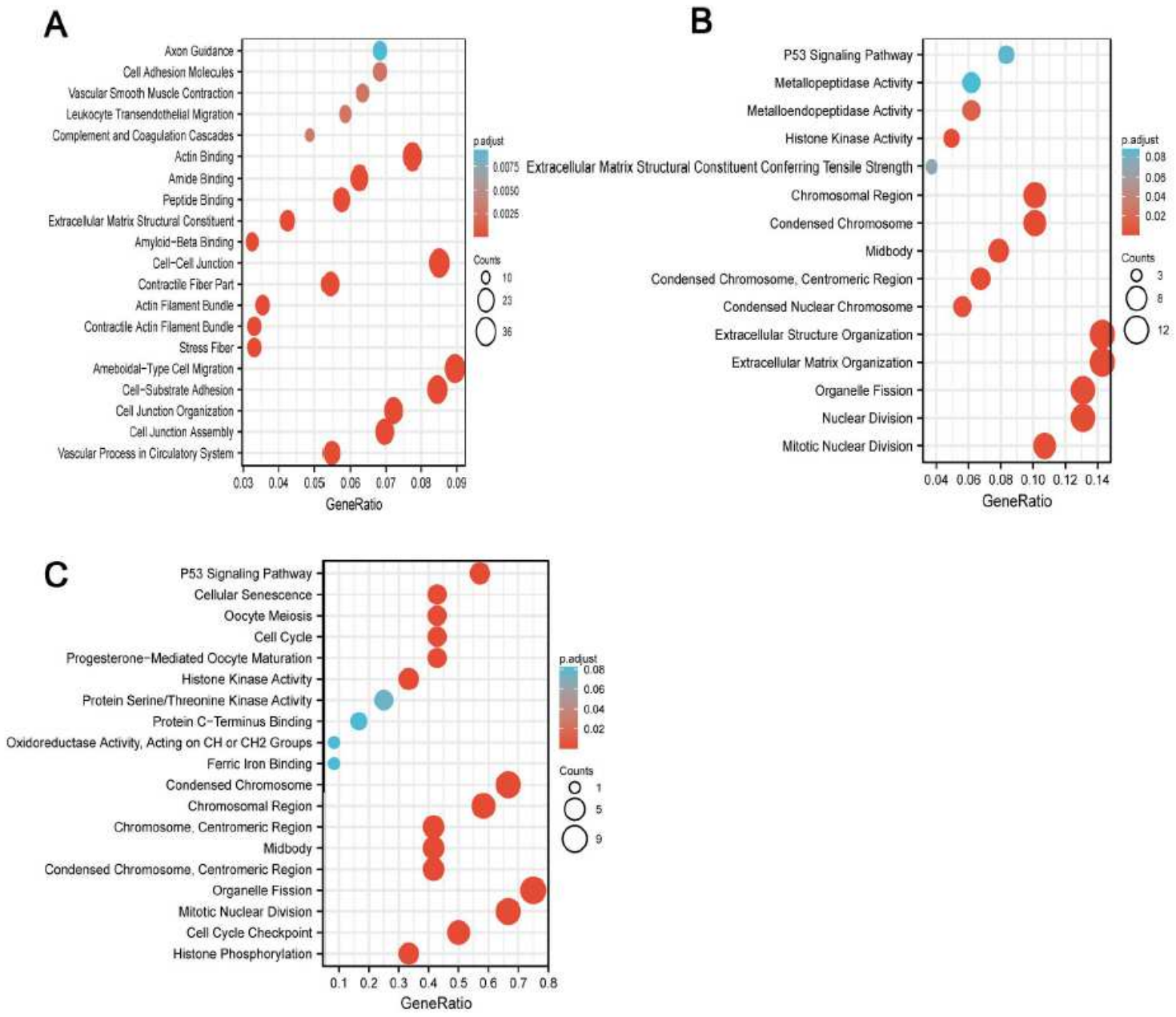

Figure 3

Download the hub gene from the DAIVED website (https://david.ncifcrf.gov/), and then use the ggplot2 package and cluster Profiler package in the R language to visualize the enrichment analysis of GO and KEGG. Panel A is the enrichment analysis of up-regulated genes, and panel B is the enrichment analysis of down-regulated genes. C is the visualization diagram of the enrichment analysis of the hub gene.

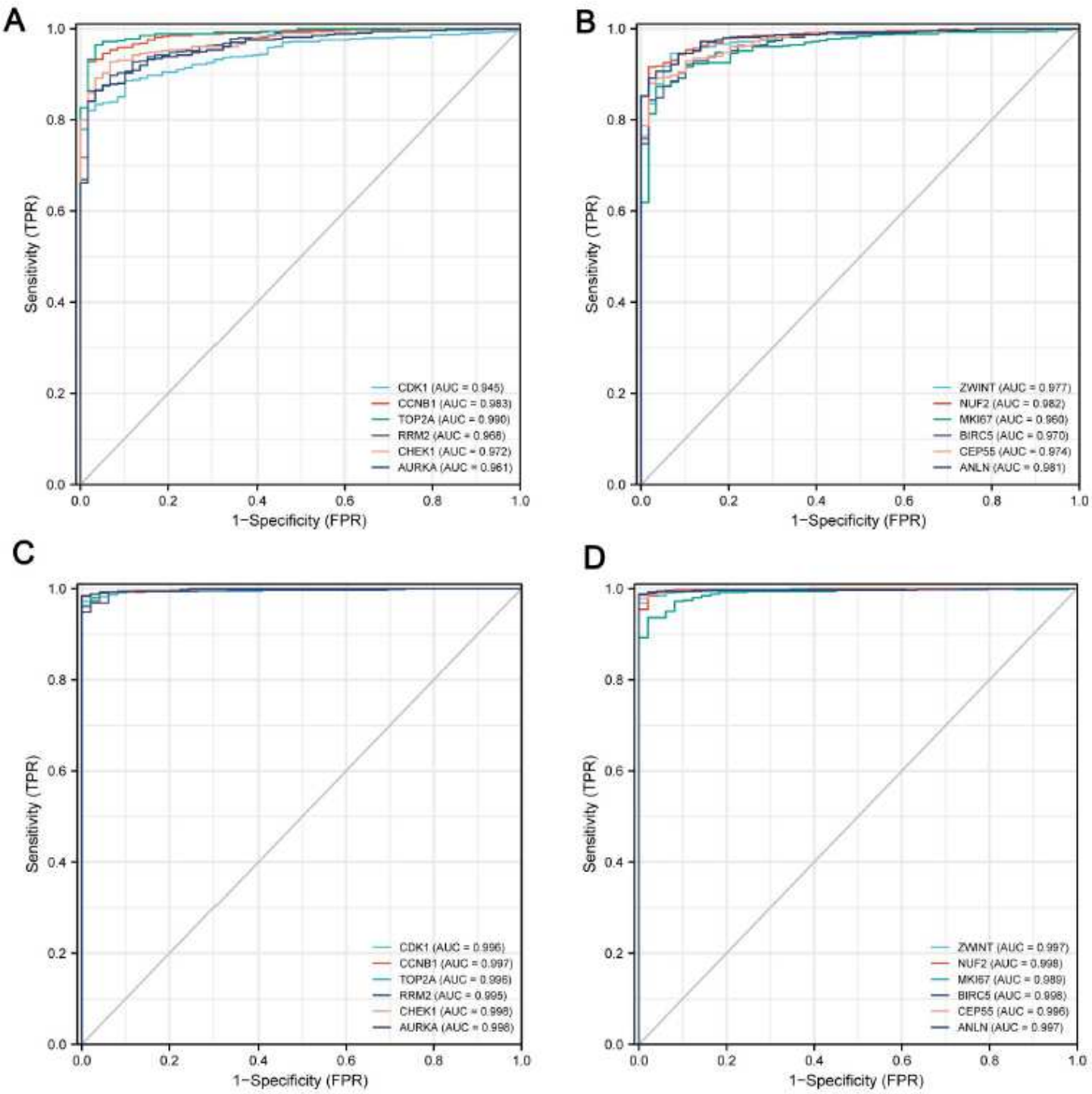

Figure 4

The 12 hub genes were analyzed with the pROC package in R language and visualized with the ggplot2 package. A and B are receiver operating characteristic curves (ROC) of LUAD, and C and D are receiver operating characteristic curves (ROC) of LUSC.

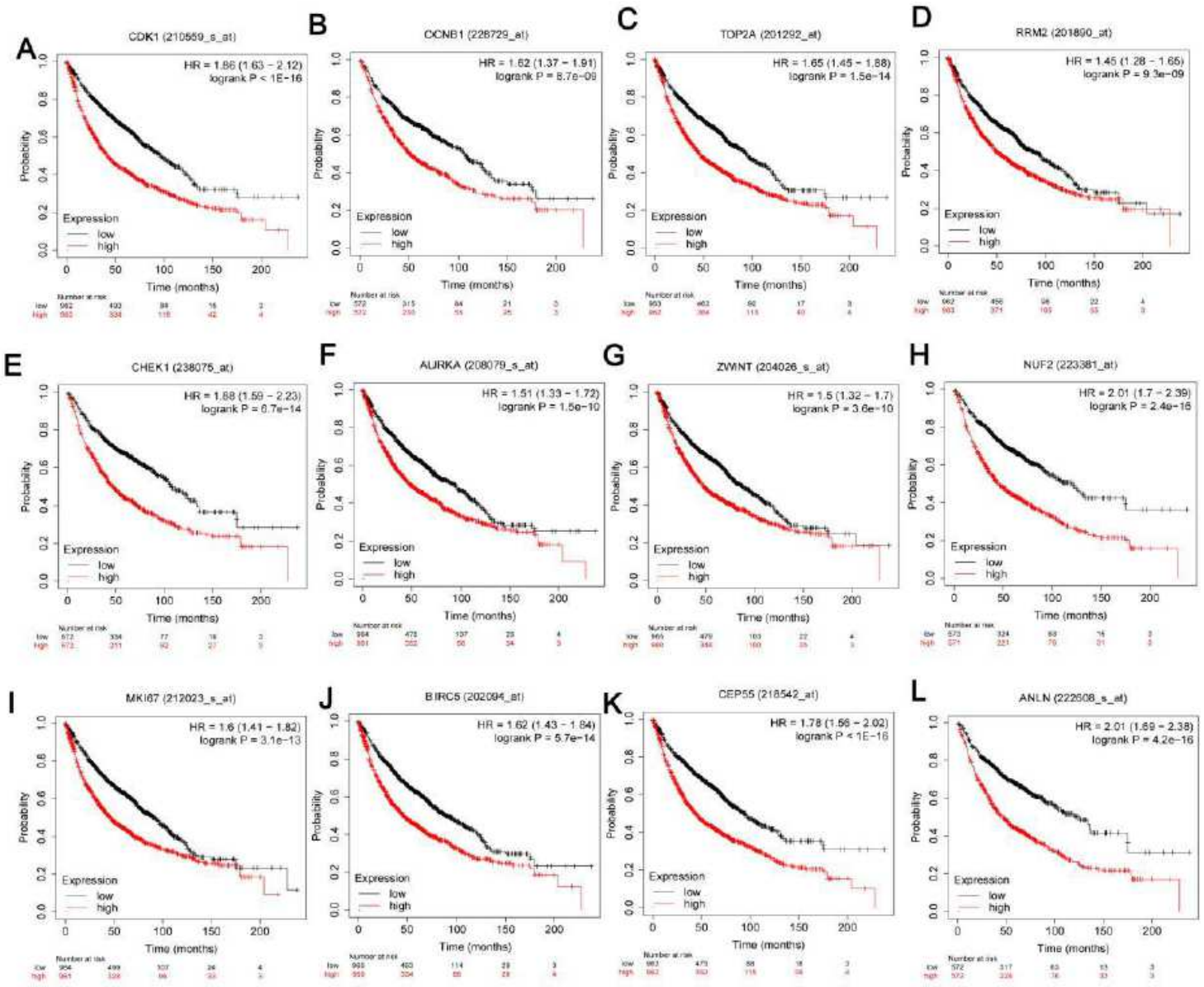

Figure 5

Overall survival analyses of 12 hub genes were performed using Kaplan-Meier Plotter online platform. P<0.05 was considered statistically significant.

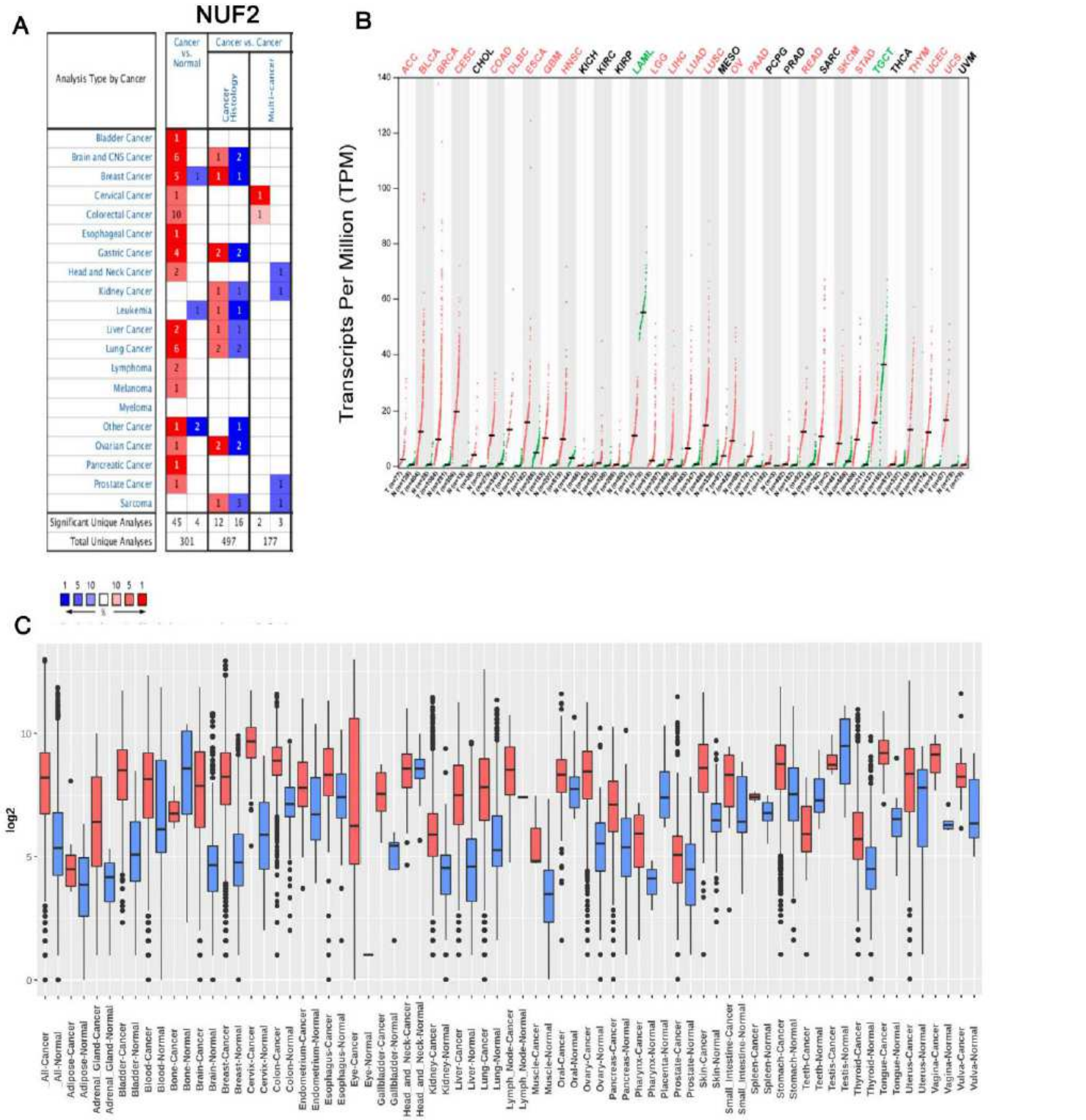

Figure 6

NUF2 mRNA expression in a variety of cancer types: (a) Comparison shows The datasets of NUF2 mRNA overexpression (left column, red) and underexpression (right column, blue) are in cancer and normal tissues. The graphic representation is derived from the Oncomine database(Available from https://www.oncomine.org/resource/login.html), and the threshold is to use the following parameters: p value is 1E-4, fold change is 2, and gene ranking is 10%. (B) Expression of NUF2 expression in 33 human

cancers (Gene Expression Profiling Interactive Analysis 2) (URL https://gepia2.cancer-pku.cn) obtained from the Cancer Genome Atlas via GEPIA2: dot map The gene expression profiles in all tumor samples and paired normal tissues are shown. Each point represents the expression of the sample. (C) The expression pattern of NUF2 mRNA in tumors and tissues. The corresponding normal tissues: The data on the expression of NUF2 mRNA in various types of cancer is retrieved from the GENT (gene expression in normal and tumor tissue) database (available at http://medicalgenomics.kribb.re.kr/GENT/). The boxes represent the median and the 25th and 75th percentiles. Dots represent outliers. The red box represents tumor tissue, and the blue box represents normal tissue.

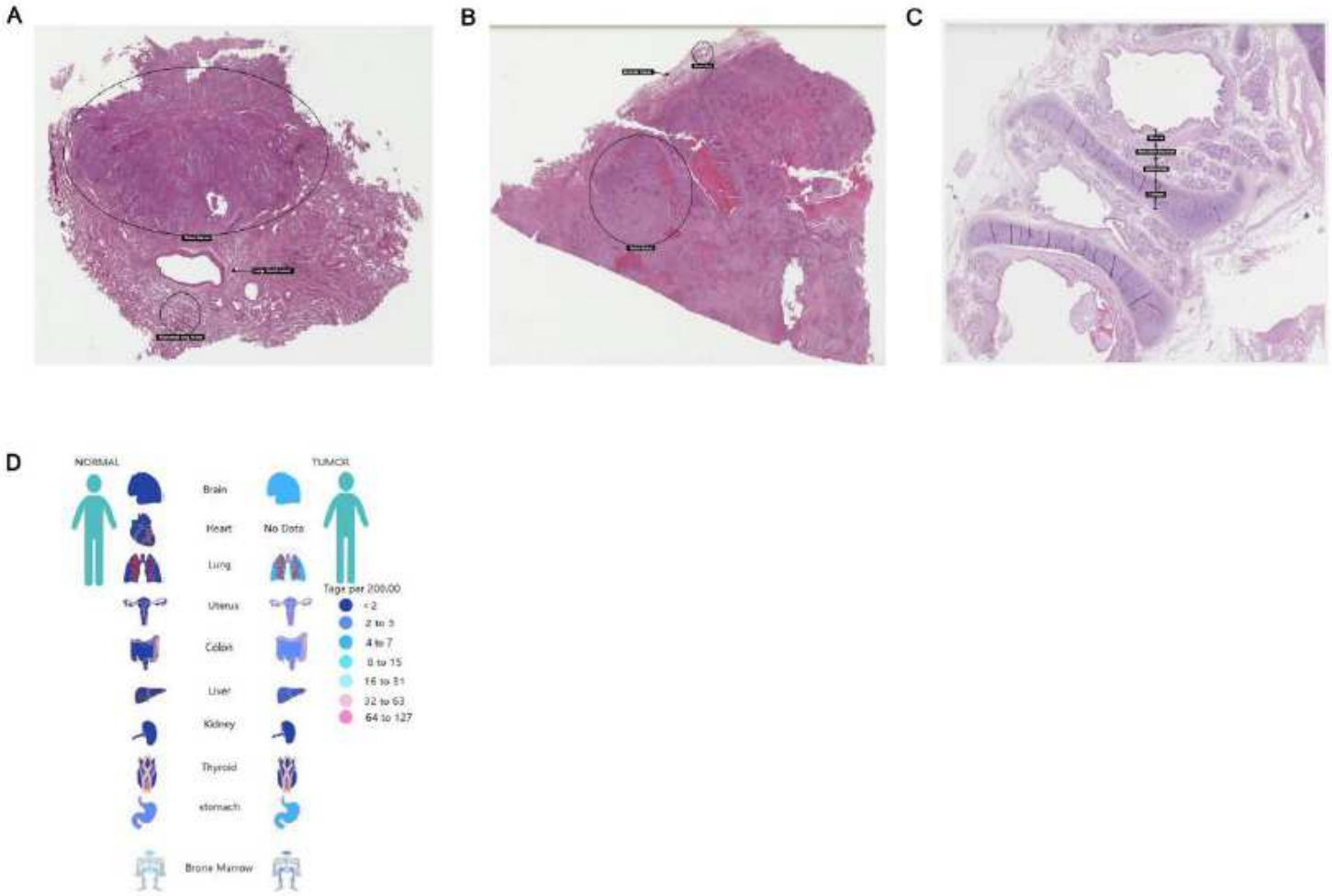

Figure 7

Use proteinatlas network database (https://www.proteinatlas.org/) to analyze: A is the HE staining result of NUF2 in lung adenocarcinoma, and B is the HE staining result of NUF2 in lung squamous cell carcinoma. C is the HE staining result of NUF2 in normal bronchus. D is the use of SAGE to analyze NUF2 in human cancers.

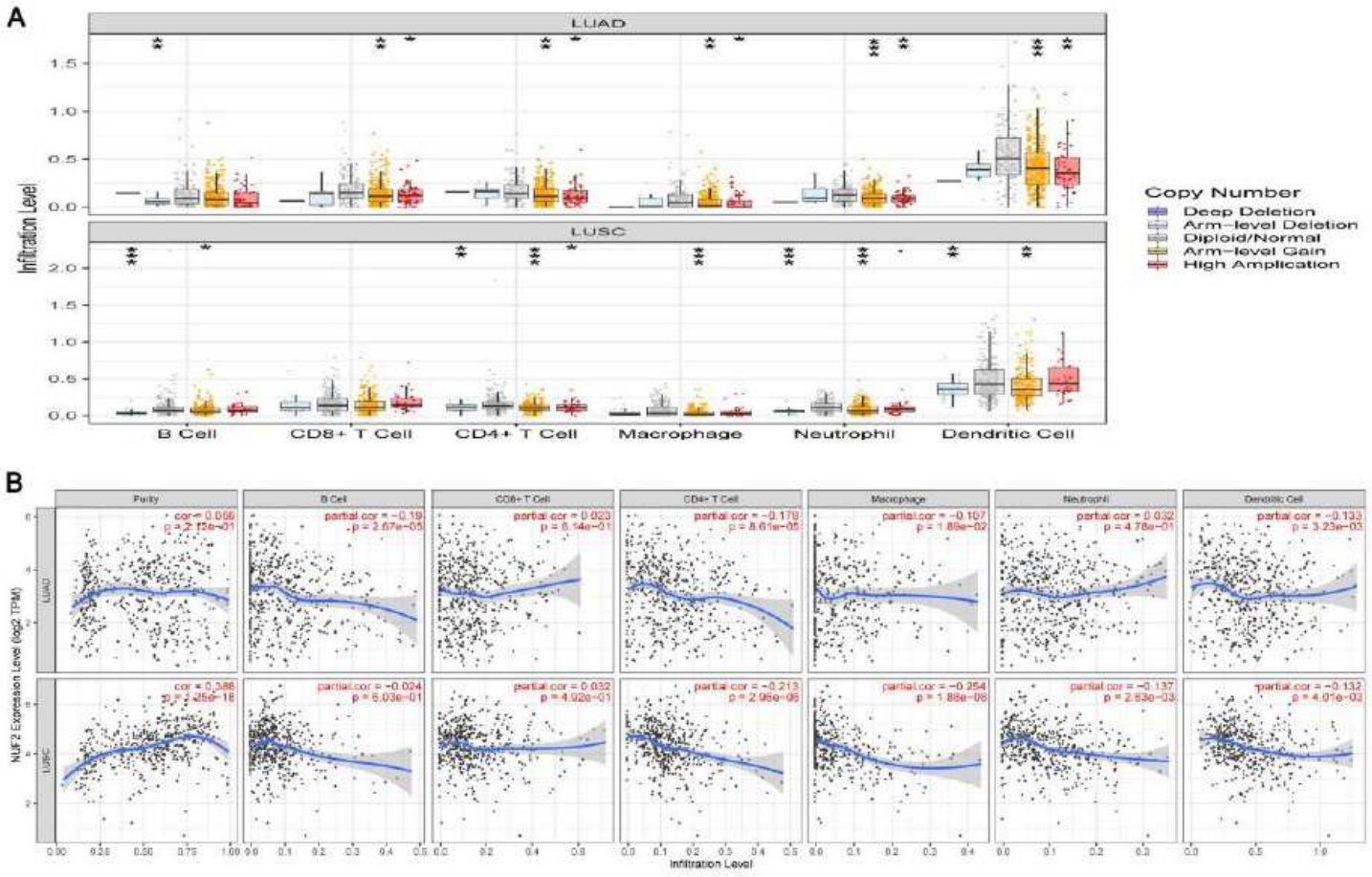

Figure 8

（A）Use the TIMER website (https://cistrome.shinyapps.io/timer/) show To explore the correlation between tumor immune cell infiltration and somatic cell copy number of NUF2 in LUAD and LUSD respectively. According to the copy number of NUF2, samples are divided into five categories (Deep Deletion, Arm-level Deletion, Diploid/Normal, Arm-level Gain, High Amplication), compare the distribution of immune cells infiltrated among the five types of samples. P-Value significant Codes: 0≤ ***（0.001≤**（0.01≤*（0.05.（B）Use the TIMER website (https://cistrome.shinyapps.io/timer/) to view the correlation between immune cell and tumor purity and NUF2 expression in LUAD and LUSD in the TCGA database.

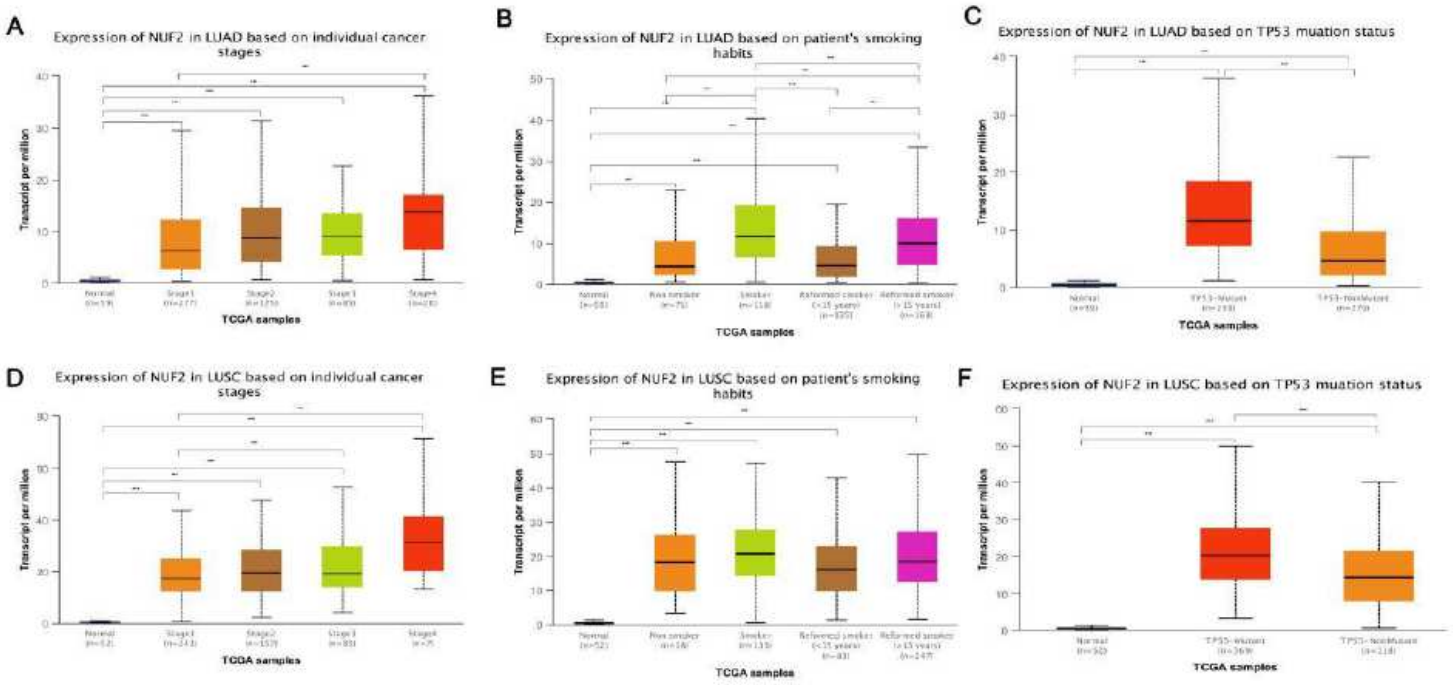

Figure 9

(A-C) Use the UALCAN website (http://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl) respectively shows the expression differences of NUF2 lung adenocarcinoma tumors in different grades, normal population and lung adenocarcinoma The expression of NUF2 in different smokers and the difference in the expression of NUF2 in TP53 mutant and non-mutated lung adenocarcinoma. (D-F) The differences in the expression of NUF2 lung squamous cell carcinoma tumors of different grades, the expression of NUF2 in normal people and different smokers of lung squamous cell carcinoma, and the difference in expression of NUF2 in TP53 mutated and non-mutated lung squamous cell carcinoma. *** indicates that the difference is statistically significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tables.pdf