Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Screening prognostic markers for non-small cell lung cancer based on data mining and bioinformatics analysis

Bin Han
Guangdong Pharmaceutical University https://orcid.org/0000-0002-5851-9425
Kaushik Chandra Aman
Jiangnan University
Dongqing Wei
Shanghai Jiao Tong University
Shulin Zhang
Shanghai Jiao Tong University School of Medicine
Minjie Meng (mminjie@126.com)
https://orcid.org/0000-0002-5169-6808

Research

Keywords: Non-small cell lung cancer, GEO database, PPI network, Cytoscape, Kaplan-Meier plotter

Posted Date: May 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-27707/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background

At present, non-small cell lung cancer has a high morbidity and mortality, and the recurrence and metastasis situation is serious. It is impossible to accurately predict the prognosis of cancer patients clinically. Biomarker is a kind of biomolecule with wide application prospects, and its potential in cancer prognosis is gradually revealed, and it is expected to be applied clinically.

Results

We integrated four gene expression profiles (GSE19188, GSE19804, GSE101929 and GSE18842) from the GEO database and screened the commonly differentially expressed genes using the GEO2R online tool. We screened 952 commonly differentially expressed genes. Gene ontology analysis showed that CDEGs were mainly enriched in biological processes such as cell adherin, angiogenesis and positive regulation of angiogenesis, and KEGG pathways such as ECM-receptor interaction and cell adherin molecules (CAMs). Up-regulation of G2 and S phase-expressed protein 1(GTSE1) expression is associated with poor prognosis of lung adenocarcinoma(LADE) and lung squamous cell carcinoma(LUSC). Up-regulation of Neuromedin-U(NMU) expression, down-regulation of Proto-oncogene c-Fos(FOS) and Cyclin-dependent kinase inhibitor 1C(CDKN1C) is only associated with poor prognosis of LADE.

Conclusions

We believe that GTSE1, NMU, FOS, and CDKN1C have potential application value as prognostic markers for lung adenocarcinoma, and are of great significance for lung adeno carcinoma efficacy evaluation and relapse monitoring. At the same time, GTSE1 may also be used as a new target for cancer treatment New ways.

Background

As one of the malignant tumors, lung cancer has the highest morbidity and mortality worldwide[1]. According to different pathological types, lung cancer is mainly divided into small cell lung cancer(SCLC) and non-small cell lung cancer(NSCLC), of which non-small cell lung cancer accounts for 85% of all lung cancer[2]. The clinical treatment methods for lung cancer mainly include surgical resection, chemotherapy, physical radiation therapy, targeted therapy and immunotherapy. Only 20–30% of patients with NSCLC are suitable for surgical resection. The toxic and side effects of chemotherapy and physical radiation therapy significantly limit its clinical application prospects. Targeted therapy is currently a widely used clinical treatment, but secondary drug resistance is a challenging problem in targeted therapy. Immunotherapy shows the advantages of significantly enhanced efficacy, reduced side effects and long-lasting effects, but the immune effect due to changes in the tumor microenvironment has reduced the therapeutic effect. Mining potential prognostic markers or therapeutic targets can further understand and understand the direction and mechanism of tumor development, and provide patients

with personalized treatment plans, monitor treatment effects, and improve the quality of life of lung cancer patients.

At present, the combination of omics technology with high throughput and bioinformatics analysis is still an important and efficient research method in medical research in order to discover target molecules related to diseases. In addition, it is also a reliable method to bioinformatics analysis for combination of a large amount of omics data to find targets with potential application value. For example, studies on oral cancer[3], glioma[4], colorectal cancer[5], osteosarcoma[6], lung cancer[7–12] and ovarian cancer[13]. Through this research method, multiple prognostic markers for NSCLC such as KIAA1522, PHLPP2, and TOP2A have been discovered, and subsequent studies have further proved that abnormally high expression of KIAA1522 is an independent factor affecting the poor prognosis of NSCLC, and is also related to platinum resistance[8]. Reduced expression of PHLPP2 in lung cancer can predict lung cancer progression[9]. TOP2A and TPX2 can jointly regulate the development of lung cancer, TOP2A may be a prognostic marker for lung cancer[7]. Due to the strong heterogeneity, high malignancy, and recurrence and metastasis of lung cancer, a single marker is difficult to meet the complex and changing clinical needs. However, the combination of markers can complete the prognosis relatively comprehensively, highlighting the advantages of stable accuracy. In this study, we screened four key differentially expressed genes of NSCLC based on Meta analysis and assessed their prognostic value.

Materials And Methods

Microassay information

Gene expression profiles(GSE19188[14], GSE19804[15], GSE101929[16] and GSE18842[17]) were obtained from the GEO database. Microarray data of GSE19188, GSE19804, GSE101929 and GSE18842 were all on account of GPL570 Platforms ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) which included 54 NSCLC tissues and 49 matched normal lung tissues, 60 NSCLC tissues and 60 matched normal lung tissues, 30 NSCLC tissues and 34 matched normal lung tissues and 46 NSCLC tissues and 45 matched normal lung tissues, respectively(Table 1).

Table 1
Information of four non-small cell lung cancer data sets from GEO database.

Data set	Normal sample	Cancer sample	Total
GSE 19188	49	54	103
GSE 19804	60	60	120
GSE 101929	34	30	64
GSE 18842	45	46	91

Identification and function enrichment of DEGs

Differentially expression gene of each series was analyzed by GEO2R online web(https://www.ncbi.nlm.nih.gov/geo/geo2r/). The criteria were set as adjusted P value < 0.01 and |logFC| > 1. The first step was to screen DEGs in each data set. The second step was to use the online tool(http://bioinformatics.psb.ugent.be/webtools/Venn/) to create a venn map to screen for commonly differentially expressed genes(CDEGs) in the four data sets. The function enrichment and KEGG(KEGG: Kyoto Encyclopedia of Genes and Genomes) pathway enrichment were performed by the Database for Annotation, Visualization and Integrated Discovery(DAVID, https://david.ncifcrf.gov/home.jsp), and the function enrichment including biological process(BP), cellular component(CC) and molecular function(MF). The criteria were set as P value < 0.01.

PPI Network construction and visualization

Protein protein interaction network of CDEGs was performed by the Search Tool for Retrieval of Interacting Genes(STRING, https://string-db.org/cgi/), which is a free tool for researchers. The minimum required interaction score set as 0.04. Further visualization of the PPI network using Cytoscape version 3.7.2 and highlighting closely related modules via the MCODE plugin. Select the highest-scoring protein in each module for more in-depth analysis.

Kaplan-Meier survival analysis

The prognostic value of CDEGs in patients with non-small cell lung cancer was analyzed by Kaplan-Meier Plotter online database(https://kmplot.com/analysis/index.php?p=background), which combines data from GEO, EGA(European Genome-phenome Archive) and TCGA(The Cancer genome Atlas) databases and is able to assess survival-related 54K genes in 21 tumors. In this study, we used NSCLC information from the database to analyze the prognostic value of CDEGs. P value < 0.05 indicates statistical significance.

Validation of CDEGs expression levels and correlation analysis

For the screening of promising CDEGs, we verified their expression levels in lung adenocarcinoma samples through the GEPIA(Gene Expression Profiling Interactive Analysis) online database(http://gepia.cancer-pku.cn/index.html). The cutoff values were set as |logFC| > 1 and P < 0.01. In addition, we also assessed the correlation between expression levels and clinical stage of tumors and investigated whether valuable CDEGs are independent influencing factors influencing the prognosis of NSCLC.

Results And Discussion

Screening of CDEGs

We collected four NSCLC datasets(GSE101929, GSE18842, GSE19188 and GSE19804) from the GEO database and performed differential expression screening. In these four data sets, there are 3179, 3162,

2601 and 1404 differentially expressed genes, of which 952 genes were commonly differentially expression genes(CDEGs), including 256 up-regulated genes and 696 down-regulated genes(Fig. 1).

Enrichment of function and KEGG pathway

To get an overview of the role and participation of 952 CDEGs in the development of NSCLC, we performed a gene ontology analysis using the DAVID online tool. We selected the five most significant terms of enrichment in each GO classification enrichment result(Table 2). The analysis showed that biological process significantly associated with the development of non-small cell lung cancer include cell adherin, extracellular matrix organization, angiogenesis, positive regulation of angiogenesis and collagen catabolic process. In addition, cell components such as proteinaceous extracellular matrix, extracellular space, extracellular region, extracellular matrix and collagen trimer and molecular function such as heparin binding, integrin binding, protein binding, calcium ion binding and metalloendopeptidase activity were also closely with the development of non-small cell lung cancer. The results of KEGG pathway enrichment shown that Malaria, Cell cycle, ECM-receptor interaction, Cell adhesion molecules (CAMs) and Protein digestion and absorption signaling pathways were most closely related to the occurrence of non-small cell lung cancer.

Table 2
Enrichment of function and KEGG pathway of commonly differentially expression genes in non-small cell lung cancer. BP, biological process. CC, cellular component. MF, molecular function. FDR, false discovery rate.

Category	Term	Gene count	P Value	FDR
GOTERM_BP	cell adhesion	61	1.89E-12	3.47E-09
GOTERM_BP	extracellular matrix organization	35	7.54E-11	1.38E-07
GOTERM_BP	angiogenesis	36	6.76E-10	1.24E-06
GOTERM_BP	positive regulation of angiogenesis	23	2.71E-08	4.97E-05
GOTERM_BP	collagen catabolic process	16	2.82E-07	5.17E-04
GOTERM_CC	proteinaceous extracellular matrix	50	1.50E-16	1.55E-13
GOTERM_CC	extracellular space	122	2.23E-12	3.18E-09
GOTERM_CC	extracellular region	138	3.51E-12	5.00E-09
GOTERM_CC	extracellular matrix	43	1.50E-10	2.14E-07
GOTERM_CC	collagen trimer	22	1.24E-09	1.77E-06
GOTERM_MF	heparin binding	29	1.41E-09	2.22E-06
GOTERM_MF	integrin binding	23	2.55E-09	4.01E-06
GOTERM_MF	protein binding	479	5.79E-07	9.11E-04
GOTERM_MF	calcium ion binding	59	3.40E-05	0.053497
GOTERM_MF	metalloendopeptidase activity	17	7.19E-05	0.113069
KEGG_Pathway	Malaria	15	1.71E-07	2.22E-04
KEGG_Pathway	Cell cycle	20	3.24E-05	0.042042
KEGG_Pathway	ECM-receptor interaction	16	5.50E-05	0.071385
KEGG_Pathway	Cell adhesion molecules	20	2.10E-04	0.272536
KEGG_Pathway	Protein digestion and absorption	15	2.36E-04	0.306116

PPI Network analysis and visualization

We performed a PPI network analysis using the STRING online tool to facilitate systematic understanding of the interactions between CDEGs and to identify key genes involved in NSCLC development. In order to more clearly and intuitively recognize the key genes in the network, we conducted a visual analysis through Cytoscape software and modularized the network through the MCODE plugin. Four modules were selected for further analysis in 28 modules. Among the four modules, there are 69, 26, 37 and 61 nodes and 2125, 173, 248 and 214 edges(Table 3), respectively, indicating that these genes are very closely

related and have multiple interacting molecules. We selected the highest-scoring gene in each module for further study in order to be able to screen out the most representative target molecules for each module, in turn GTSE1, NMU, FOS and CDKN1C (Fig. 2).

Table 3
Information of four modules from PPI network. A, B, C and D represent different modules respectively. CDEGs, commonly differentially expression genes.

Module	Node	Edge	CDEGs
A	69	2125	MCM2, ECT2, MKI67, UBE2C, TTK, NDC80, AURKA, RMI2, KIF20A, FEN1, KIF14, NUF2, MAD2L1, PTTG1, KIF11, CCNB1, TK1, CENPU, CDCA7, SHCBP1, CCNE2, NUSAP1, CDKN3, CDC6, CDK1, CCNB2, MND1, DEPDC1, GTSE1, ANLN, CHEK1, BIRC5, KIF4A, CENPF, CEP55, UBE2T, KIF15, GINS2, TYMS, EZH2, TOP2A, ASPM, DLGAP5, ZWINT, E2F8, STIL, CDC20, MELK, CASC5, CDCA3, PBK, CCNA2, PRC1, GMNN, HMMR, FOXM1, KIAA0101, KIF2C, TPX2, NCAPG, NEK2, BUB1, DEPDC1B, UHRF1, RRM2,MCM4, CENPK, BUB1B, FANCI
B	26	173	SSTR1, FPR1, EDN1, CXCR2 ,PECAM1, ANXA1, ANGPT1, GNG11, TLR8, P2RY13, NMU, KDR, CXCL2, SELE, S1PR1, ACE, CX3CR1, CSF3, VWF, PPBP, CXCL13, CXCL3, TEK, CDH5, P2RY14, AGTR2
C	37	248	ICAM1, COL6A6, FOS, COL1A2, IL1B, IRF1, CLDN5, CAT, MMP1, PTX3, PTGS2, COL4A3, SOCS3, AGTR1, CCL2, FGF2, RUNX2, COL10A1, THY1, COL5A2, COL1A1, SELP, COL11A1, NES, PPARG, LEPREL1, PLA2G1B, MMP9, THBD, TLR4, CAV1, SPP1, COL5A1, COL13A1, PLAU, BMP2, PLOD2
D	61	214	CXCL12, HECW2, CD93, CLEC12A ,FZD4, NEDD4L, CD274, CYR61, DAMTS1, KLF2, CD36, GPC3, CDKN1C, GOLM1, GRK5, LMO7, MCEMP1, PTGER4, ARRB1, IL6, CD69, ITGA1, CHRDL1, ADAMTSL4, CALCRL, ADAMTS9, IL33, SLC2A5, ALOX5, FS1, SBSPON, FBXO32, RNF182, CTTN, PTPRB, CFP, S100A4, ADRB1, ACTR3, SEMA5A, ADAMTS12, SH3GL3, RXFP1, AGER, ZBTB16, THBS2, PRF1, CP, RAMP3, DUSP1, MMP13, HBEGF, RNF144B, ADAMTS8, VIPR1, LDLR, RAMP2, ATF3, ADAMTSL3, SPARCL1, RAP1A

Survival analysis

We selected four key genes, of which GTSE1(logFC = 1.32, adjusted P value < 0.001) and NMU(logFC = 2.81, adjusted P value < 0.001) were up-regulated in tissues of patients with non-small cell lung cancer, and the expression levels of FOS(logFC = -2.22, adjusted P value < 0.001) and CDKN1C(logFC = -1.56, adjusted P value < 0.001) were down-regulated. To assess the prognostic value of GTSE1, NMU, FOS, and CDKN1C in patients with NSCLC, we analyzed another group of NSCLC cases from GEO, EGA and TCGA databases using Kaplan-Meier Plotter. The results showed that high expression levels of GTSE1(P value < 0.01) and NMU(P value < 0.01) were closely related to the shorter overall survival of LADE patients, with statistical significance. In contrast, high expression levels of FOS(P value < 0.01) and CDKN1C(P value < 0.01) were significantly associated with longer overall survival in patients with LADE, with statistically significant significance(Fig. 3). However, except for NMU, FOS and CDKN1C, only GTSE1 expression levels are associated with patient prognosis in LUSC. These findings demonstrate that GTSE1 and NMU

high expression, low expression of FOS and CDKN1C can be used as indicators of poor prognosis of LADE.

GEPIA analysis

We have demonstrated the potential clinical value of GTSE1, NMU, FOS and CDKN1C in the prognosis of patients with NSCLC. To confirm whether the expression levels of GTSE1, NMU, FOS and CDKN1C in the LSCLC tissues were consistent with the results of the four data sets, we verified 1654 samples by GEPIA analysis. The results showed that the expression levels of GTSE1($P < 0.05$) and NMU($P < 0.05$) were significantly up-regulated, and the expression levels of FOS($P < 0.05$) and CDKN1C($P < 0.05$) were significantly down-regulated in lung adenocarcinoma and statistically significant(Fig. 4). The validation results shown that the four CDEGs were completely consistent with the results in the four data sets, further demonstrating that these four targets have good prognostic value for LADE. In addition, we conducted a correlation analysis by GEPIA to investigate whether the four targets are independent of each other affecting LADE. The results show a very weak positive or negative correlation between GTSE1 and NMU, GTSE1 and FOS, FOS and CDKN1C, respectively(Fig. 5). The relationship between gene expression levels and clinical stage showed that the expression levels of NMU, FOS and CDKN1C were not related to the clinical stage of lung adenocarcinoma except GTSE1(Fig. 6). The expression level of GTSE1 increases with the tumor stage, indicating that it can be used as a potential indicator for determining the development of tumors. Based on all the above findings, GTSE1, NMU, FOS and CDKN1C have good prognostic value for patients with lung adenocarcinoma.

The important reason for the high mortality rate of lung cancer is that lung cancer is easy to metastasize and relapse during the treatment. Therefore, it is urgent to solve the clinical problem by accurately predicting potential prognostic markers of tumor progression status. Transcriptionomics with high-throughput advantages can provide powerful help for researchers in medical research to facilitate screening of target molecules. Therefore, we integrated four mRNA expression profiles of non-small cell lung cancer derived from the GEO database and studied them.

By comparing NSCLC tissues with paired paracancerous tissues, we screened 952 CDEGs from four expression profiles, including 256 CDEGs with up-regulated expression and 696 CDEGs with down-regulated expression. Gene ontology analysis showed that CDEGs were mainly enriched in biological processes such as cell adherin, angiogenesis and positive regulation of angiogenesis, and KEGG pathways such as ECM-receptor interaction and cell adherin molecules (CAMs). Similarly, Piao JJ et al. also reported this result[11].

We selected GTSE1, NMU, FOS and CDKN1C for further research through PPI network analysis because they are the core genes. It has been reported that GTSE1 is highly expressed in tumors such as liver cancer and melanoma, and is associated with poor prognosis of patients[18, 19]. Moreover, GTSE1 may be involved in tumorigenesis and progression by regulating p53 phosphorylation[20, 21]. In another liver cancer study, it was also proved that the down-regulation of GTSE1 has a good effect of promoting

apoptosis, reducing anti-apoptotic ability and enhancing the sensitivity of chemotherapeutic drugs[22]. Therefore, it can be considered that GTSE1 not only has the potential as a prognostic marker for lung adenocarcinoma, but may also be used as a new target to provide a more effective targeted drug delivery route for cancer treatment. NMU is well known for its uterine smooth muscle contraction inducer. Meanwhile, It also participates in the formation and development of various tumors. For example, Koji Takahashi et al. reported that the positive rate of NMU in NSCLC and SCLC was as high as 68% and 82%, and the overexpression of NMU was verified at protein level and transcriptional level[23]. In addition, studies have shown that overexpression of NMU is also produced in HER2 overexpressing breast cancer, and overexpression of NMU in breast cancer is associated with poor prognosis in patients[24, 25]. Not only that, but similar reports have been reported in the study of Clear cell renal cell carcinoma and endometrial carcinoma[26–28]. CDKN1C is a tumor suppressor gene, which is down-regulated in studies related to gastric cancer[29], bladder cancer[30], pancreatic cancer[31], lung cancer[32] and breast cancer[33], and low expression levels are associated with poor prognosis in patients. Importantly, all of the above research results strongly support our analysis results. In addition, GTSE1, NMU, FOS and CDKN1C have no correlation with each other, indicating that each target can be used alone as a prognostic marker for NSCLC. In conclusion, we believe that GTSE1, NMU, FOS and CDKN1C can be used as potential markers for the prognosis of lung adenocarcinoma, and provide a basis for clinical lung adenocarcinoma efficacy evaluation and recurrence monitoring. At the same time, GTSE1 may also be used as a new target for cancer treatment.

Conclusion

We performed a differential analysis of large-scale lung small cell lung cancer samples and matched paracancerous tissues based on bioinformatics, and selected four core CDEGs for in-depth study. Then through meta analysis, expression level verification and correlation analysis, we believe that GTSE1, NMU, FOS and CDKN1C have potential and clinical application value as prognostic markers of LADE. At the same time, GTSE1 may also be used as a new target for cancer treatment.

Declarations

Acknowledgements

Thanks to Kaushik Chandra Aman, Dongqing Wei, Shulin Zhang and Minjie Meng for their indispensable contributions to the paper.

Authors' contributions

Conceived and designed the ideas: Minjie Meng, Shulin Zhang. Analyzed the data: Bin Han, Kaushik Chandra Aman, Dongqing Wei. Wrote the paper: Bin Han. All authors read and approved the final manuscript.

Funding

This work was supported by National Mega-Project of China(2017ZX10201301-003-001), National Natural Science Fundation of China(81871613) and Science and Technology Plan Projects of Guangdong Province (Grant Number 2016A050502064).

Ethics approval and consent to participate

Because the human data of our manuscript is all from TCGA public data, we believe that we do not need additional ethics approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All available URLs and online tools have been shown in the text.

References

1. Siegel RL, Miller KD, et al. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7–34.
2. Siegel RL, Miller KD, et al. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30.
3. Luca F, Gabriella L, et al. Identification of Novel MicroRNAs and Their Diagnostic and Prognostic Significance in Oral Cancer. Cancers (Basel), 2019, 11(5).
4. Di J, Shenglan L, et al. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging. 2019;10(4):592–605.
5. Pan JH, Zhou H, et al. LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. Front Immunol. 2019;10:6.
6. Hu G, Cheng Z, et al. Identification of potential key genes associated with osteosarcoma based on integrated bioinformatics analyses. J Cell Biochem. 2019;120(8):13554–61.
7. Ma W, Wang B, et al. Prognostic significance of TOP2A in non-small cell lung cancer revealed by bioinformatic analysis. Cancer Cell Int. 2019;19:239.
8. Liu YZ, Yang H, et al. KIAA1522 is a novel prognostic biomarker in patients with non-small cell lung cancer. Sci Rep. 2016;6:24786.
9. Wang H, Gu R, et al. PHLPP2 as a novel metastatic and prognostic biomarker in non-small cell lung cancer patients. Thoracic cancer. 2019;10(11):2124–32.
10. Huang H, Huang Q, et al. Differentially Expressed Gene Screening, Biological Function Enrichment, and Correlation with Prognosis in Non-Small Cell Lung Cancer. Med Sci Monit. 2019;25:4333–41.

11. Piao J, Sun J, et al. Target gene screening and evaluation of prognostic values in non-small cell lung cancers by bioinformatics analysis. Gene. 2018;647:306–11.
12. Ni M, Liu X, et al. Identification of Candidate Biomarkers Correlated With the Pathogenesis and Prognosis of Non-small Cell Lung Cancer via Integrated Bioinformatics Analysis. Frontier in Genetics, 2018, 9.
13. Feng H, Gu ZY, et al. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. Journal of ovarian research. 2019;12(1):35.
14. Hou J, Aerts J, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. 2010;5(4):e10312.
15. Lu TP, Tsai MH, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. Cancer Epidemiol Biomarkers Prev. 2010;19(10):2590–7.
16. Mitchell KA, Zingone A, et al. Comparative Transcriptome Profiling Reveals Coding and Noncoding RNA Differences in NSCLC from African Americans and European Americans. Clin Cancer Res. 2017;23(23):7412–25.
17. Sanchez-Palencia A, Gomez-Morales M, et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. Int J Cancer. 2011;129(2):355–64.
18. Wu X, Wang H, et al. GTSE1 promotes cell migration and invasion by regulating EMT in hepatocellular carcinoma and is associated with poor prognosis. Sci Rep. 2017;7(1):5129.
19. Xu T, Ma M, et al. High G2 and S-phase expressed 1 expression promotes acral melanoma progression and correlates with poor clinical prognosis. Cancer Sci. 2018;109(6):1787–98.
20. Liu A, Zeng S, et al. Overexpression of G2 and S phase-expressed-1 contributes to cell proliferation, migration, and invasion via regulating p53/FoxM1/CCNB1 pathway and predicts poor prognosis in bladder cancer. Int J Biol Macromol. 2019;123:322–34.
21. Liu XS, Li H, et al. Polo-like kinase 1 phosphorylation of G2 and S-phase-expressed 1 protein is essential for p53 inactivation during G2 checkpoint recovery. EMBO Rep. 2010;11(8):626–32.
22. Ren F, Zhao T. Silencing GTSE 1 gene expression induces apoptosis of hepatocarcinoma cells through p53 pathway. Tumor. 2020;40:113–21.
23. Takahashi K, Furukawa C, et al. The neuromedin U-growth hormone secretagogue receptor 1b/neurotensin receptor 1 oncogenic signaling pathway as a therapeutic target for lung cancer. Cancer Res. 2006;66(19):9408–19.
24. Shetzline SE, Rallapalli R, et al. Neuromedin U: a Myb-regulated autocrine growth factor for human myeloid leukemias. Blood. 2004;104(6):1833–40.
25. Wu Y, McRoberts K, et al. Neuromedin U is regulated by the metastasis suppressor RhoGDI2 and is a novel promoter of tumor formation, lung metastasis and cancer cachexia. Oncogene. 2007;26(5):765–73.
26. Ketterer K, Kong B, et al. Neuromedin U is overexpressed in pancreatic cancer and increases invasiveness via the hepatocyte growth factor c-Met pathway. Cancer Lett. 2009;277(1):72–81.

27. Przygodzka P, Papiewska-Pajak I, et al. Neuromedin U is upregulated by Snail at early stages of EMT in HT29 colon cancer cells. Biochim Biophys Acta. 2016;1860(11 Pt A):2445–53.
28. Zhang S, Wang Q, et al. Identification and analysis of genes associated with papillary thyroid carcinoma by bioinformatics methods. Biosci Rep, 2019, 39(4).
29. Shin JY, Kim HS, et al. Mutation and expression of the p27KIP1 and p57KIP2 genes in human gastric cancer. Exp Mol Med. 2000;32(2):79–83.
30. Oya M, Schulz WA. Decreased expression of p57(KIP2)mRNA in human bladder cancer. Br J Cancer. 2000;83(5):626–31.
31. Sato N, Matsubayashi H, et al. Epigenetic down-regulation of CDKN1C/p57KIP2 in pancreatic ductal neoplasms identified by gene expression profiling. Clin Cancer Res. 2005;11(13):4681–8.
32. Sun Y, Jin SD, et al. Long non-coding RNA LUCAT1 is associated with poor prognosis in human non-small lung cancer and regulates cell proliferation via epigenetically repressing p21 and p57 expression. Oncotarget. 2017;8(17):28297–311.
33. Qiu Z, Li Y, et al. Downregulated CDKN1C/p57(kip2) drives tumorigenesis and associates with poor overall survival in breast cancer. Biochem Biophys Res Commun. 2018;497(1):187–93.

Figures

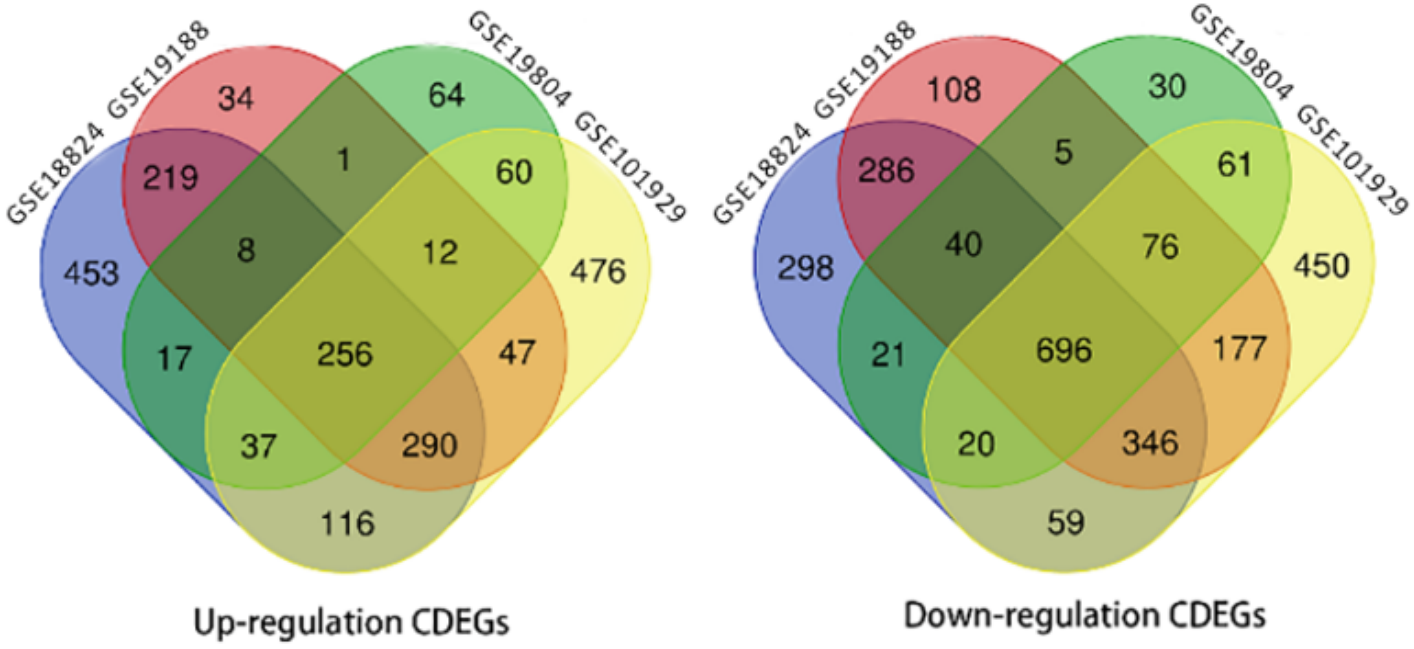

Figure 1

CDEGs from four data sets. Analysis of the four data sets revealed that there were 256 commonly up-regulated differentially expressed genes and 696 commonly down-regulated differentially expressed

genes, respectively.

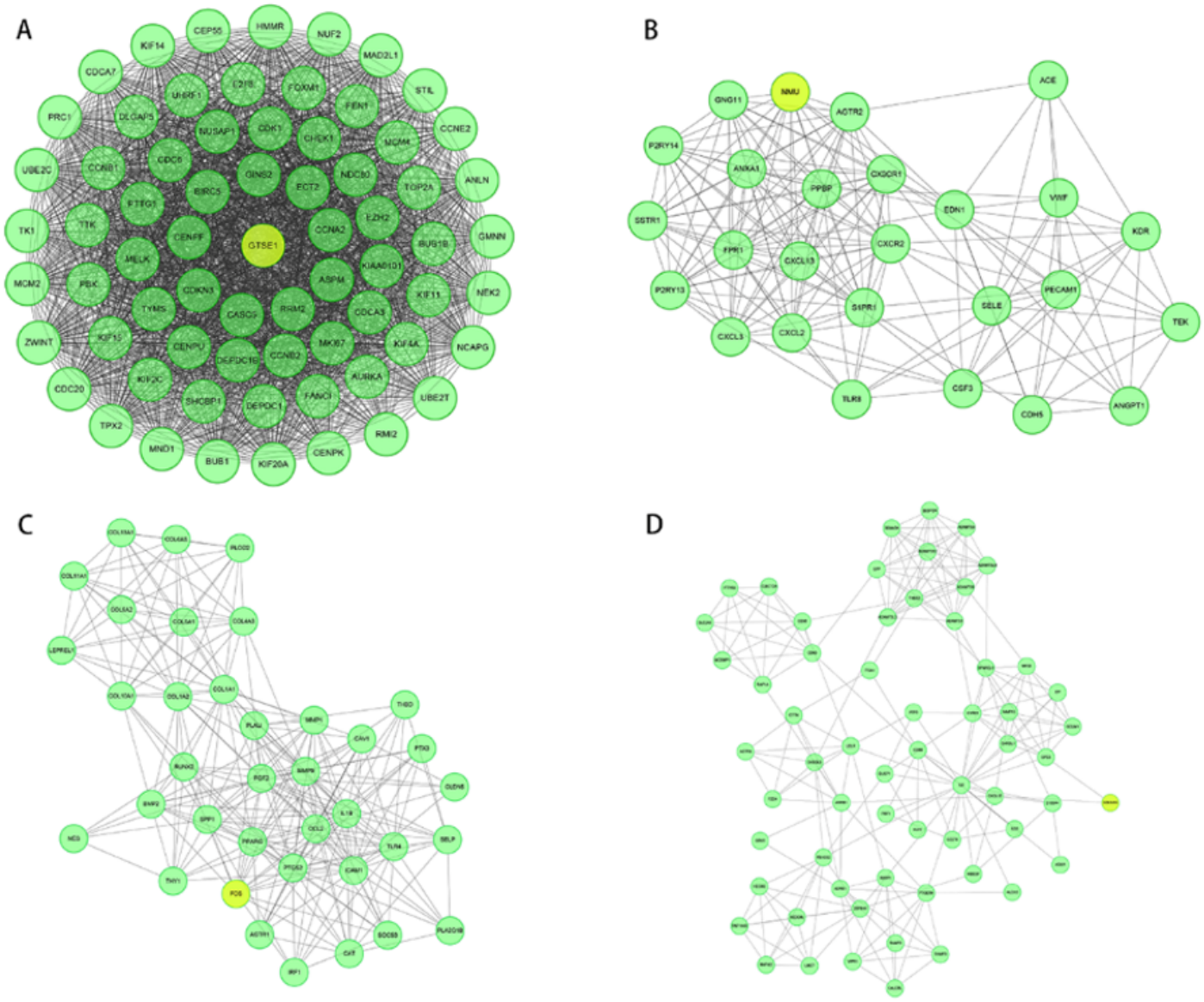

Figure 2

PPI network visualization of CDEGs in NSCLC. Visual analysis of CDEGs through Cytoscape software and MCODE plugin after PPI network analysis using STRING. A. Module A contains 69 nodes and 2125 edges, and the highest-scoring gene is GTSE1. B. Module B contains 26 nodes and 173 edges, and the highest-scoring gene is NMU. C. Module C contains 37 nodes and 248 edges, and the highest-scoring gene is FOS. D. Module D contains 61 nodes and 214 edges, and the highest-scoring gene is CDKN1C. Yellow represents the highest-scoring gene.

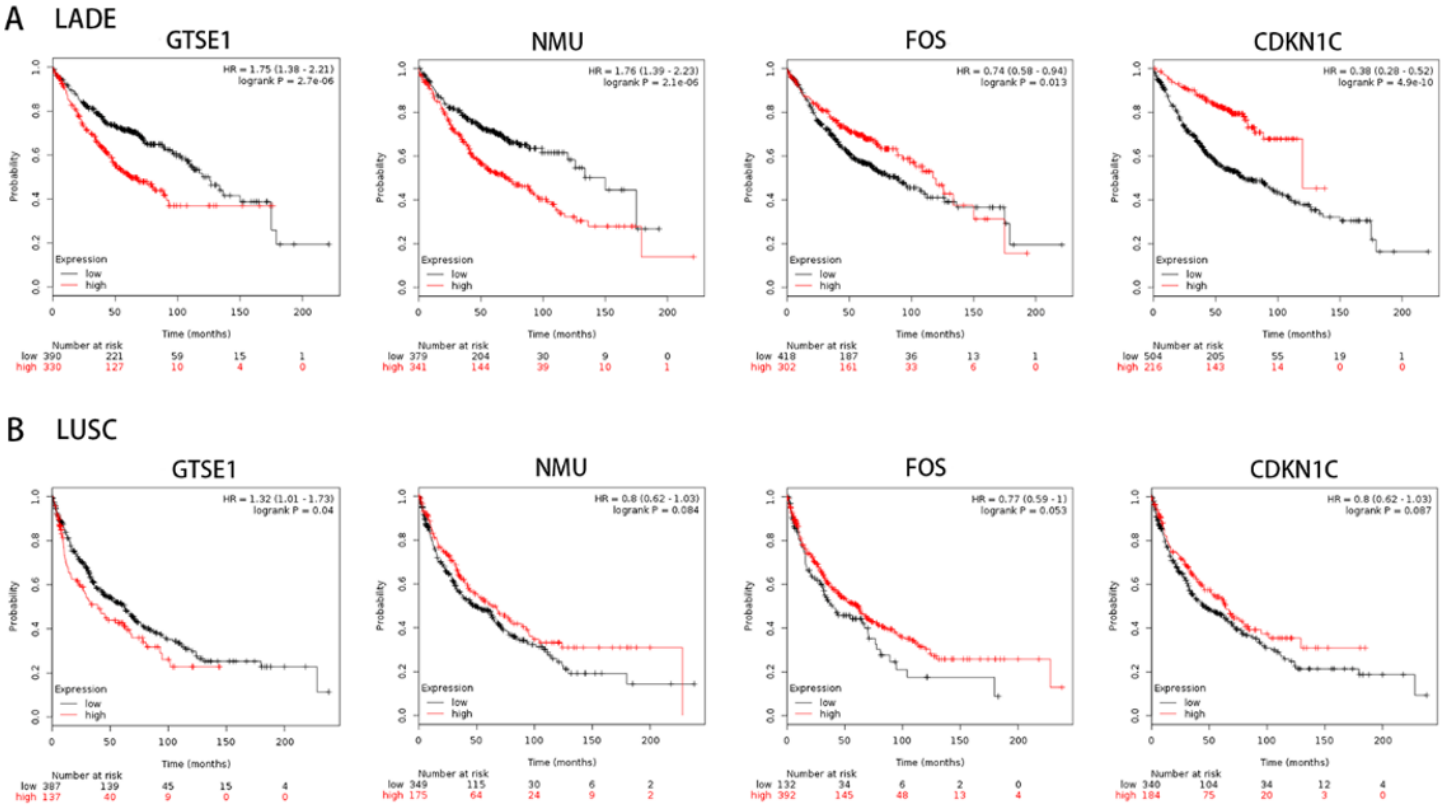

Figure 3

Prognosis analysis of CDEGs in patients with lung adenocarcinoma(LADE) and lung squamous cell carcinoma(LUSC). A indicates high expression of GTSE1 and NMU in lung adenocarcinoma patients, and low expression of FOS and CDKN1C is associated with poor prognosis. B indicates that high expression of GTSE1 in lung squamous cell carcinoma patients is associated with poor prognosis, while NMU, FOS and CDKN1C are not statistically significant.

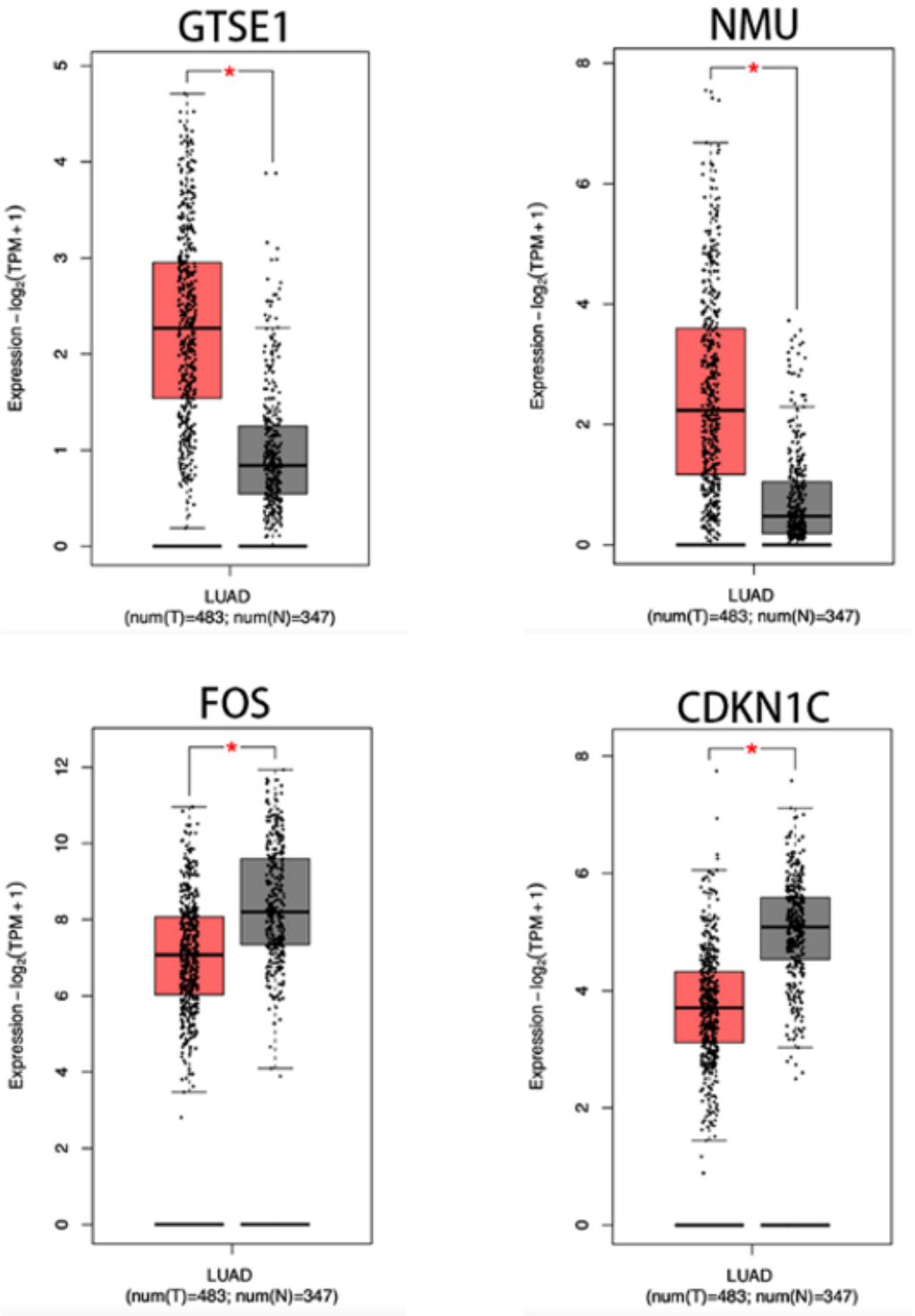

Figure 4

Verification of expression levels of GTSE1, NMU, FOS and CDKN1C in lung adenocarcinoma tissues. Validation by GEPIA showed significantly higher expression of GTSE1 and NMU in lung adenocarcinoma compared with normal tissues, and significantly lower expression of FOS and CDKN1C. Asterisks means statistially significant.

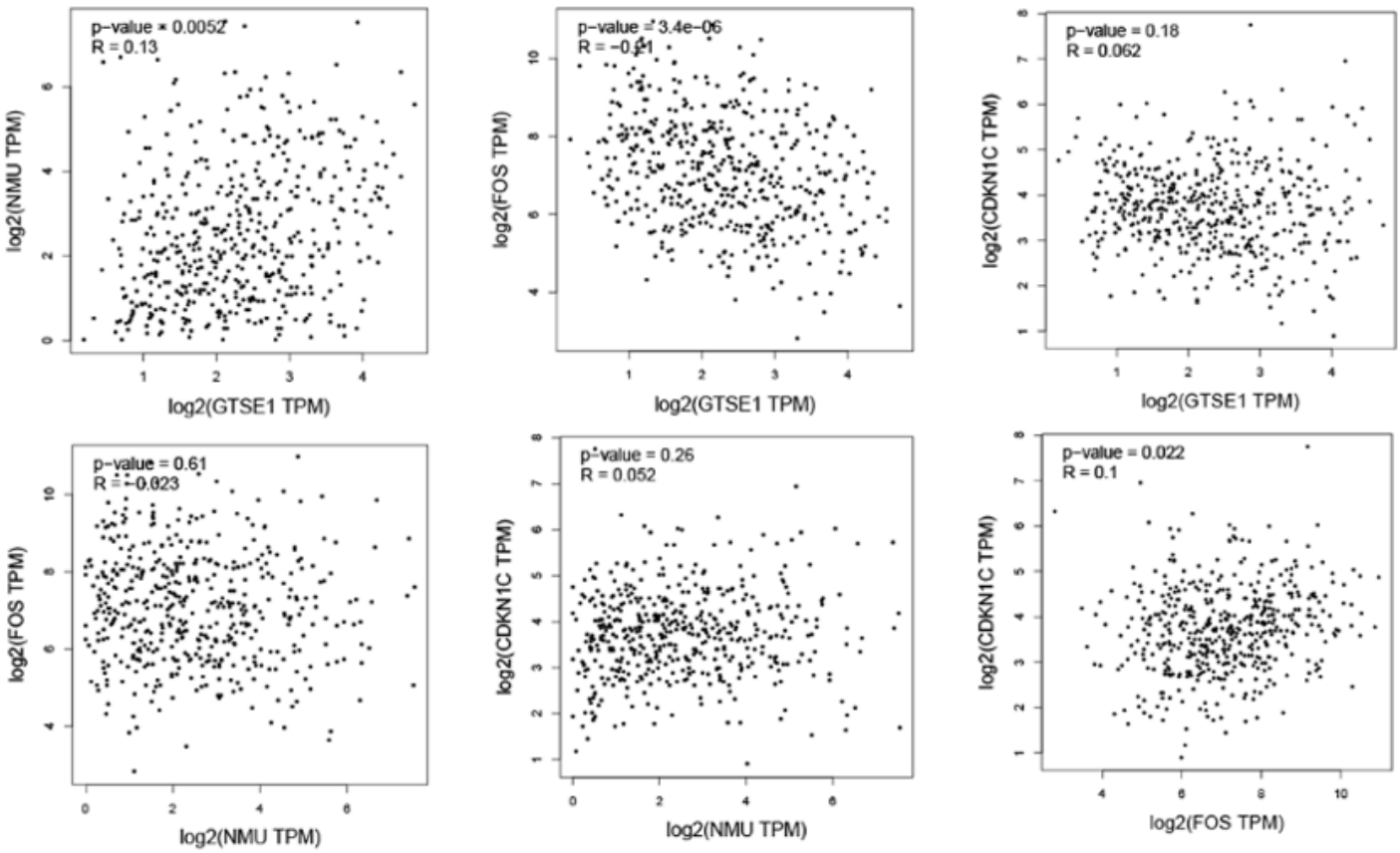

Figure 5

Correlation analysis of GTSE1, NMU, FOS and CDKN1C expression levels in lung adenocarcinoma tissues. The results show a weak positive correlation between GTSE1 and NMU, FOS and CDKN1C, and a weak negative correlation between GTSE1 and FOS. Correlation analysis using the Pearson rank sum test, $P < 0.05$ means statistically significant.

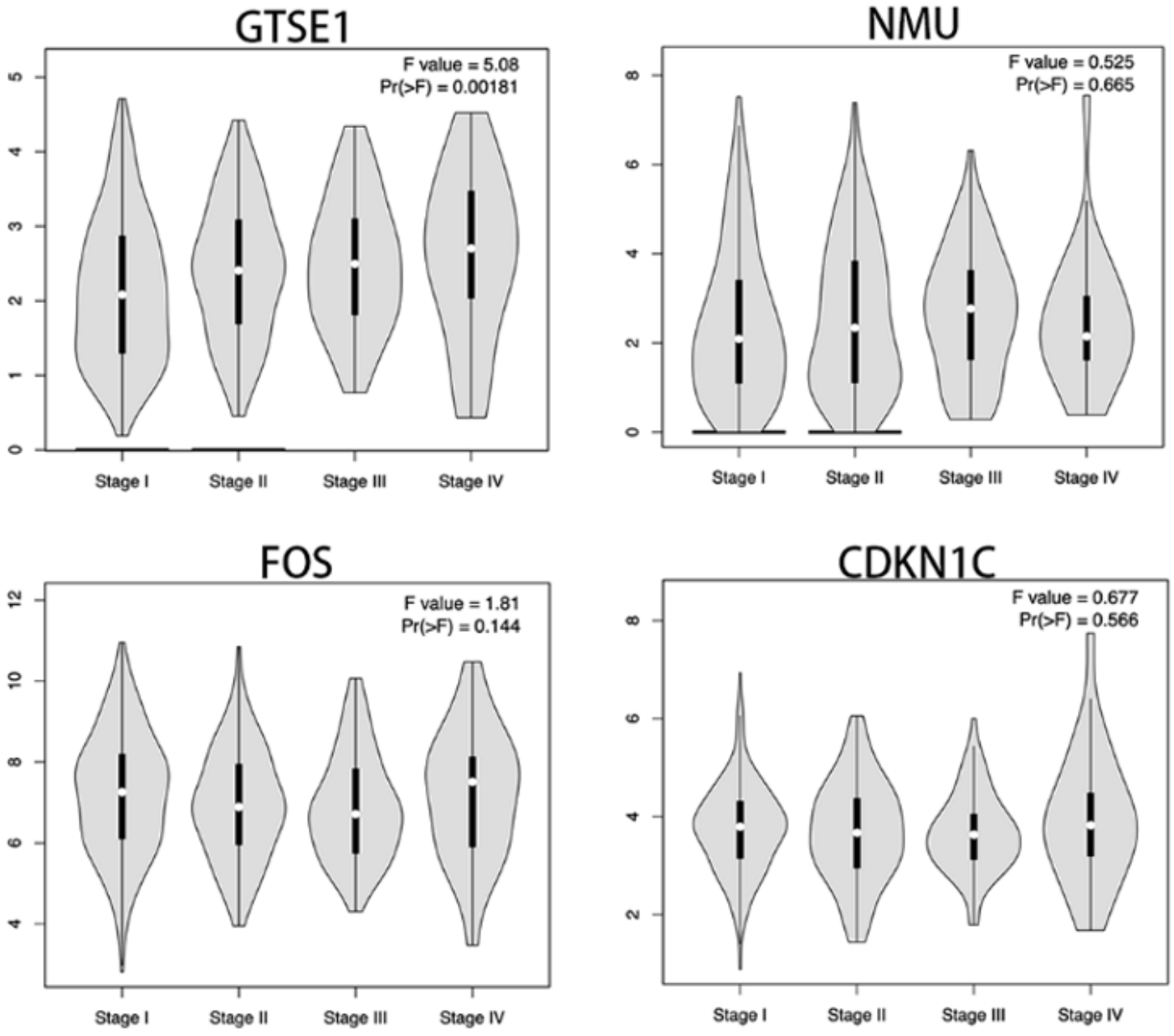

Figure 6

Analysis of the correlation between GTSE1, NMU, FOS and CDKN1C expression levels and clinical stage of lung adenocarcinoma. Except for NMU, FOS and CDKN1C, only the expression level of GTSE1 is correlated with lung adenocarcinoma staging and is positively correlated.